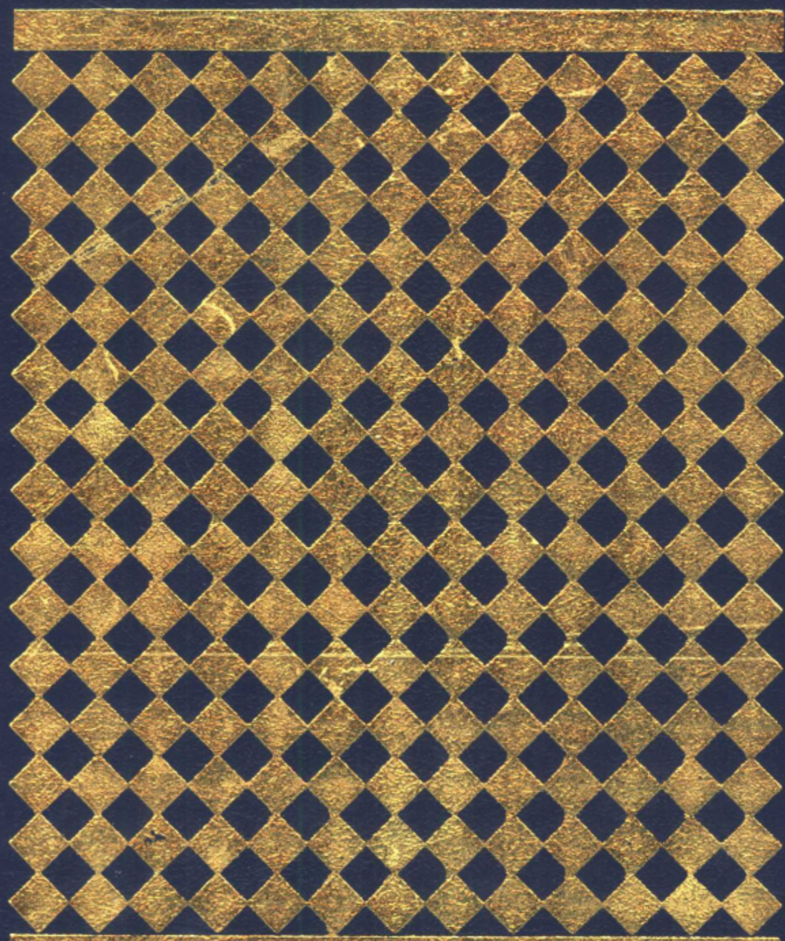
概率论基础及其应用

王梓坤 著



北京师范大学出版社

概率论基础及其应用

王 梓 坤

北 京 师 范 大 学 出 版 社

图书在版编目(CIP)数据

概率论基础及其应用/王梓坤著．—北京：北京师范大学出版社，1996.3
ISBN 7—303—03632—6

Ⅰ．概… Ⅱ．王… Ⅲ．概率论 Ⅳ．0211

中国版本图书馆 CIP 数据核字(96)第 01745 号

概率论基础及其应用
王 梓 坤
北京师范大学出版社出版发行
(邮编 100875 北京新街口外大街 19 号)
北京师范大学印刷厂印刷 全国新华书店经销
开本：787×1092 1/16 印张：22.75 字数：576 千
1996 年 5 月第 1 版 1996 年 5 月第 1 次印刷
印数：1—1 500 册
ISBN7—303—03632—6/O·183 定价：36.00 元

序

本书的目的是想比较严格地叙述概率论的基础知识，并介绍它的一些应用。总共八章，前三章是核心部分，其中包括概率论的基本概念和定理；第四章是随机过程的初步导引；后面四章分别介绍概率论在数理统计、随机模拟、计算方法以及可靠性问题中的一些简单应用。

随着科学和社会的进步，概率论获得了新的生命力。理论一与实际相结合，便立即呈现出“忽如一夜春风来，千树万树梨花开”的繁荣景象。目前，概率论已广泛应用到许多实际部门中，要在一本书内叙述各方面的应用是困难的。因此，本书所谓的应用，只不过是一些方面的一些应用。

在写作过程中，我们注意了下列两点：

1. 为了使叙述比较清楚，我们广泛使用了“概率空间”的概念。有了它才能明确地给出一些最基本的定义，如事件、随机变数等等，否则，恐怕很难回答象“随机变数的函数是否仍是随机变数”、“独立随机变数 ξ_1,ξ_2 的函数 $g(\xi_1),g(\xi_2)$ 是否仍独立”等这样一些无法避免的问题。但另一方面，由于引进了概率空间，就必须用到一些最基本的测度论知识，这似乎不属本书的范围。不过我们发现，如果只考虑概率论基础的需要，那么用到的测度论并不太多，在本书中完全可以自给自足，而且所用的篇幅很少。这样做还可以多少填平一些普通概率论与较高深的概率论分支之间的鸿沟。这种叙述方式也许还是一种新尝试。

2. 对于基本概念，我们希望能讲明它们的背景和应用。在实践中所体会到的直观形象有助于抓住本质，它常常是理论的先导，并为理论提供思路、模型与方法；严格的逻辑证明和计算有时无非是直觉的一种数学加工和精确化而已。虽然如此，严谨的数学论证仍是十分重要的。

书中列举了较多的例题，其中有些是著名的问题，它们在概率论的发展上起过一定的作用。前五章附有习题及详细解答，有些习题也可当作例题看，以利于自学。

§1.4 (三) 表示第一章第四节第三段。凡标有星号 (*) 的章、节、段、定理或证明，都可在初次阅读时略去，待必要时补看。

本书大部分内容曾在南开大学数学系多次讲授过。芦桂章、孙瑊二位老师为本书提供了习题及解答；吴荣老师详细审阅了全部底稿，并提出许多改进意见，作者对他们表示衷心的感谢。

由于水平所限，书中一定会有不少的错误和缺点，恳请批评指正。

作者　1976 年

第二版 前言

本书初版于 1976 年，后连印刷三次。承一些高等院校、科研单位用作教本或参考书。不少热心的读者提出了许多中肯的批评，甚至寄来勘误表；这些意见在这一版中已全被采纳，使本书为之生色。作者对他们的帮助表示最真诚的感谢。作者还衷心地感激韩丽娟、洪良辰、廖昭懋、蒋铎、洪吉昌等各位教授，他们关心本书的再版，并给于了许多帮助，第二版书末增加一附录《论随机性》，作为概率哲学的初探。

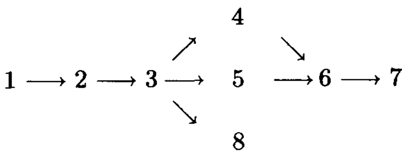
作者　1992 年 10 月

内 容 简 介

本书叙述概率论的基本知识和它的一些应用。总共八章，前三章介绍概率论，第四章是随机过程的初步导引，后四章则分别介绍概率论在数理统计、随机模拟、计算方法和可靠性问题中的一些应用。读者对象为高等院校理工科学员、教师及科学工作者。

目　　录

各章间关系图



第一章 事件与概率

§1.1 概率论的现实背景

(一) 概率论的研究对象. 自然界有许多现象，我们完全可以预言它们在一定条件下是否会出现. 例如："同性的电互相排斥","在标准大气压下，水加热到 100°C 时必定沸腾"等等是一定会出现的，而上述现象的反面，即"同性的电互相吸引","在标准大气压下，水加热到 100°C 时不沸腾"等等是必然不会出现的. 在一定条件下必然出现的现象叫必然事件；在一定条件下必然不出现的现象叫不可能事件. 必然事件的反面是不可能事件.

然而自然界还有许多现象，它们在一定条件下可能出现也可能不出现，这种现象称为随机事件，或简称为事件. 例如"掷伍分钱硬币得正面","明年北京七月间的平均温度是 28° C"等等都是随机事件.

概率论是数学的一个分支，它研究的对象是随机事件的数量规律性.

我们常常通过随机试验来观察随机事件. 例如：事件"正面"是随机试验"掷硬币"的一个可能的结果，而事件"28°C"是随机试验"观察明年北京七月间的平均温度"的一个可能的结果.

一般地，设 E 为一试验，如果不能事先准确地预言它的结果，而且在相同条件下可以重复进行，就称为随机试验，以 ω 表它的一个可能的结果，称 ω 为 E 的一基本事件，全体基本事件的集 $\Omega=(\omega)$ 称为基本事件空间. 在具体问题中，十分重要的是：认清基本事件空间是由什么构成的，我们来举一些例子.

例1 E- 掷一枚普通的硬币而观察所出现的面 [1]；ω_1- 正， ω_2- 反，于是 Ω 由两个基本事件构成，即 $\Omega=(\omega_1,\omega_2)$.

例2 E- 自标号为 $1,2,\cdots,n$ 的 n 个同样的灯泡中任取其一， ω_i- 取得第 i 号， $\Omega=(\omega_1,\omega_2,\cdots,\omega_n)$. 这时如果简记 ω_i 为 i, 则得 $\Omega=(1,2,\cdots,n)$.

例3 E- 计算某电话交换台在上午九点钟内所得呼唤次数； ω_i- 得 i 次呼唤； $\Omega=(\omega_0,\omega_1,\omega_2,\cdots)$. 如果简记 ω_i 为 i, 则得 $\Omega=(0,1,2,\cdots)$.

例4 E- 观察北京十月间的平均温度， ω_a- 平均温度为 a°C. $\Omega=(\omega_a,-\infty<a<\infty)$. 如果简记 ω_a 为 a, 则得 $\Omega=(-\infty,\infty)$, 即全体实数集 (实际上 Ω 中有许多点是多余的，因为温度不可能低于 -273°C, 等等).

例5 E- 向平面上某块有界区域 Ω 内随意投掷质点 (例如掷球) 而观察落点位置，以 $\omega(a,b)$ 表落点的坐标为 (a,b) 的一个试验结果. 如果简记 $\omega(a,b)$ 为 (a,b), 那么基本事件空间重合于此区域 Ω.

基本事件是事件中的一种，一般的事件总是由若干个基本事件共同组成的，因而是 Ω 的一个子集. 譬如说，在例 2 中，事件 A:"取得标号不大于 3 的灯泡"是由三个基本事件

[1] 许多实际问题与此问题同类，例如：登记新生婴孩的性别， "女"或"男", 产品"合格"或"不合格", 明天天气"晴"或"非晴", 射击"中的"或"不中的"等.

"取得第 1 号""取得第 2 号""取得第 3 号"共同组成的，故 $A=(1,2,3)$. 同样，对事件 B:"取得标号为偶数的灯泡"我们有 $B=\left(2,4,6,\cdots,2\left[\frac{n}{2}\right]\right)$, 其中 $[a]$ 表示不大于 a 的最大整数. 又如在例 4 中，事件 C:"不高于 18°C"可表为 $C=(-\infty,18]$. 由此可见，每一事件对应于 Ω 的一个子集.

当且仅当事件 A 所含的一个基本事件出现时称为 A 出现.

(二) **事件的运算**. 试引进事件间的一些重要关系及对事件的运算. 以下 A,B,A_i,B_i 都表事件.

1. 如果 A 出现必导致 B 出现，就说 A 是 B 的特款，或者说 B 包含 A, 记作 $A\subset B$; 如果 $A\subset B$ 而且 $B\subset A$ 就说 A 与 B 相等，并记作 $A=B$. 例如 A: 取得第 2 号灯泡"是 B: "取得偶数号灯泡"的特款；又如 A:"平均温度不高于 18°C 是 B:"平均温度低于 20°C"的特款.

2. "二事件 A,B 中至少有一个出现"也是一事件，称此事件为 A,B 的和，记作 $A\cup B$; 显然，$A\cup B$ 也是事件："或 A 出现，或 B 出现", 类似地，事件"$A_1,\cdots,A_n$ 中至少有一出现"称为 $A_1,\cdots,A_n$ 的和，记为 $A_1\cup\cdots\cup A_n$ 或 $\bigcup\limits_{i=1}^{n}A_i$; 事件"$A_1,A_2,\cdots$ 中至少有一出现"称为可列多个事件的和，记作 $A_1\cup A_2\cup\cdots$ 或 $\bigcup\limits_{i=1}^{\infty}A_i$. 其次，"$n$ 个事件 $A_1,\cdots,A_n$, 都出现"也是一事件，称为 $A_1,\cdots,A_n$ 的交，记作 $A_1\cap\cdots\cap A_n$ 或 $A_1\cdots A_n$ 或 $\bigcap\limits_{i=1}^{n}A_i$. 类似定义 $A_1\cap A_2\cap\cdots$ 并简写为 $A_1A_2\cdots$ 或 $\bigcap\limits_{i=1}^{\infty}A_i$. 譬如说，在例 3 中，"呼唤次数为偶数"是基本事件"2''"4''"6''$\cdots$ 等的和，而"呼唤次数为偶数"与"呼唤次数为 3 的倍数"的交是事件"呼唤次数是 6 的倍数".

3. "A 出现而 B 不出现"也是一事件，称为 A 与 B 的差，记作 $A-B$. 例如："呼唤次数不小于 6"与"呼唤次数不小于 7"的差是"呼唤次数为 6"; 而"呼唤次数不小于 6"与"呼唤次数为偶数"的差是"呼唤次数为不小于 7 的奇数".

4. 不可能事件与必然事件通常也看成随机事件，分别用记号 $\emptyset$ 及 Ω 表示. 如果二事件 A,B 满足关系

$$AB=\emptyset, \tag{1}$$

也就是说，如果 A,B 不可能同时出现，就说 A,B 是互不相容的. 例如："呼唤次数大于 10"与"呼唤次数小于 10"是互不相容的.

二事件 A,B 如果满足关系

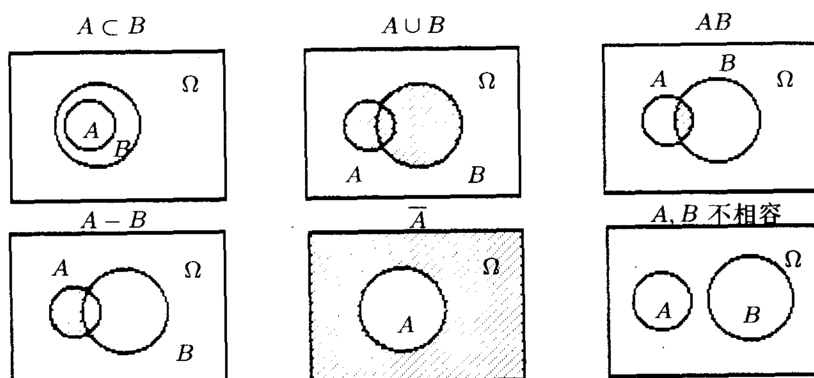
$$A\cup B=\Omega,\quad AB=\emptyset, \tag{2}$$

也就是说，A,B 中必出现其一，但 A,B 不能同时出现，就说 A 与 B 互逆，或者说 A 是 B(或 B 是 A) 的对立事件，记作 $A=\overline{B}$. 例如"呼唤次数大于 10"是"呼唤次数不大于 10"的对立事件，但却不是"呼唤次数小于 10"的对立事件.

上面说过，事件 A 可表为基本事件空间 Ω 中的一个子集，这子集仍然记为 A, 于是我们也可以从集合论的观点来看待事件，结果发现这种观点非常有用，因为上面对事件所引进的关系恰好和通常对集所引进的相应的关系一致，譬如说，设事件 A 是事件 B 的特款，于是 A 出现必导致 B 出现. 考虑含于 A 的一基本事件 ω, 当 ω 出现时，A 出现，由假

定这时 B 也出现，故此 ω 必含于 B, 从而证明了集 A,B 间有集合论中所定义的包含关系 $A\subset B$. 反之，设集 A 含于集 B, $A\subset B$; 当事件 A 出现时，必定有 A 中的某基本事件 ω 出现，既然此 $\omega\in A$, 因而 $\omega\in B$, 故事件 B 也出现，这说明 A 是 B 的特款. 由此可见，"事件 B 包含事件 A" 与 "集 B 包含集 A" 二概念是一致的. 类似地可以看到其他相应概念的一致性：事件的 "相等","和","交","差" 分别与集的 "相等","和","交","差" 一致，不可能事件 $\emptyset$ 及必然事件 Ω 分别与空集 $\emptyset$ 及全空间 Ω 一致. 从而关系式 (1),(2) 也可按集合论中意义来理解，因此，"A 是 B 的对立事件" 与 "集 A 是集 B 关于 Ω 的补集" 是一致的.

事件间的关系既然与集间相应的关系一致，我们就可以用图形来表达，如下图：



($A\cup B, AB, A-B, \overline{A}$ 分别为图中阴影部分)

(三) 概率. 当我们多次做某一随机试验 E 时，常常会察觉某些事件出现的可能性要大些，也就是说，这事件出现的次数要多些，而另一些事件出现的可能性要小些. 例如 "抽得偶数号灯泡" 的可能性大于 "抽得第 2 号灯泡" 的可能性，因为后一事件是前一事件的特款. 既然各事件出现的可能性有大有小，自然使人想到该用一个数字 $P(A)$ 来标志事件 A 出现的可能性，较大的可能性用较大的数字来标志，较小的就用较小的数字. 这数字 $P(A)$ 就称为事件 A 的概率.

然而，对于已给的事件 A, 到底应该用哪个数字来作为它的概率呢？就是说，怎样从数量上来规定 $P(A)$ 呢？这决定于所研究的随机试验 E 及 A 的特殊性，不能一概而论. 但在下列两种特殊的 (即古典型的和几何型的) 随机试验中，事件的概率容易合理地规定，并能满足实际的需要.

(A)古典型. 先考虑一个特例，即例 2. 自标号为 $1,2,\cdots,n$ 的共 n 个同样的灯泡中任取其一，如果抽取时把这些灯泡完全平等看待，不特别侧重或特别看轻某些灯泡，那么，每个灯泡被取出的可能性应该是相同的，也就是说，一切基本事件 ω_i–"取得第 i 号灯泡" 都是等可能的 (为了要实现等可能，譬如说可以这样做：在 n 张同样的纸片上各写上一个数字，从 1 到 n, 然后把纸片折好，并把次序随意搅乱，再从中任取一张，它上面的数字就代表被取中的灯泡号码). 在此例中，只有 n 个不同的基本事件.

一般地，如果随机试验 E 具有下列二性质：

1) 只有有穷多个不同的基本事件 $\omega_1,\omega_2,\cdots,\omega_n$,

2) 一切基本事件都是等可能的.

就说 E 是古典型的随机试验.

由上所述可见，例 2 及例 1 中的 E 都是古典型的.

对古典型的随机试验 $E, \Omega=(\omega_1,\omega_2,\cdots,\omega_n)$, 设事件 A 由 $k(\leqslant n)$ 个不同的基本事件组成，我们就定义 A 的概率 $P(A)$ 为

$$P(A)=\frac{k}{n}. \tag{3}$$

不可能事件 $\emptyset$ 的概率定义为

$$P(\emptyset)=0. \tag{4}$$

譬如，在例 2 中，如果设 $n=100$, 那么事件

$$A=(\text{取得偶数号灯泡})=(\omega_2,\omega_4,\cdots,\omega_{100}),\ P(A)=\frac{50}{100}=\frac{1}{2};$$
$$B=(\text{取得号数不大于 10 的灯泡})=(\omega_1,\omega_2,\cdots,\omega_{10}),\ P(B)=\frac{10}{100}=\frac{1}{10};$$
$$C=(\text{取得号数为 3 的倍数的灯泡})=(\omega_3,\omega_6,\cdots,\omega_{99}),\ P(C)=\frac{33}{100}.$$

关于古典型的定义再说几句话：那里的第二个条件要求一切基本事件都是等可能的，即等可能地出现，然而，所谓“基本事件等可能地出现”又是什么意思呢？对此我们不能下数学定义而只能做一些解释：当从问题有关的各个方面来考虑 $\omega_1,\omega_2,\cdots,\omega_n$ 时，如果它们完全处于平等的地位，谁也不比谁特别些，这时就可把它们看成是等可能的. 在目前很难用更简单的概念来定义“等可能”，就像集合论中“集合”的概念没有明确的数学定义一样. 我们无需对此感到惊异，因为如果概念甲是由较简单的概念乙来定义，那么乙也必须用更简单的概念丙来定义，如此下去，最后势必会遇到一个最基本的概念，我们再不能也不必用其它概念去定义它了，这时只需用一些例子或直观的语言去解说它的含义. “等可能”正是一个这样的基本概念.

附带指出，在实际问题中，往往只能“近似地”出现等可能，“完全的”等可能是很难见到的. 以掷硬币为例，严格说来，正、反二面也不能认为完全是等可能的. 因为两面的花纹不同，凸凹的分布不同等. 不过这些因素对出现正 (或反) 面的影响很小，因而可以把它们忽略而仍认为是等可能的.

定理1 对古典型随机试验 E, 概率具有下列性质：

(i) 对任意事件 $A, 1\geqslant P(A)\geqslant 0$;

(ii) $P(\Omega)=1$;

(iii) 设事件 $A_1,\cdots,A_m\ (m\leqslant n)$ 互不相容，则

$$P\left(\bigcup_{i=1}^{m}A_i\right)=\sum_{i=1}^{m}P(A_i). \tag{5}$$

证. 由 (3) (4) 可见 (i) 是显然的. 由于 Ω 由 n 个不同的基本事件构成，故由 (3)

$$P(\Omega)=\frac{n}{n}=1.$$

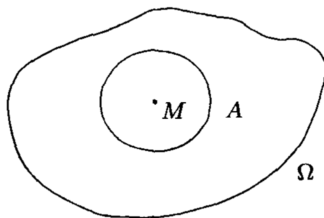
下证 (iii): 设 $A_i=(\omega_1^{(i)},\cdots,\omega_{k_i}^{(i)})$ 含 $k_i(\leqslant n)$ 个不同的基本事件；由 (3) $P(A_i)=\frac{k_i}{n}$ 显然

$$\bigcup_{i=1}^{m}A_i=(\omega_1^{(1)},\cdots,\omega_{k_1}^{(1)};\omega_1^{(2)},\cdots,\omega_{k_2}^{(2)};\cdots;\omega_1^{(m)},\cdots,\omega_{k_m}^{(m)})$$

共含 $\sum\limits_{i=1}^{m}k_i$ 个基本事件；由于 $A_1,A_2,\cdots,A_m$ 互不相容，这些基本事件是互不相同的. 因此

$$P\left(\bigcup_{i=1}^{m}A_i\right)=\frac{\sum\limits_{i=1}^{m}k_i}{n}=\sum_{i=1}^{m}\frac{k_i}{n}=\sum_{i=1}^{m}P(A_i).\ \square$$

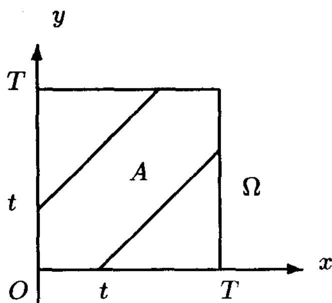
(B) **几何型**. 几何型随机试验是例 5 中 E 的一般化与精确化. 设 Ω 是 n 维空间中的勒贝格 (Lebesgue) 可测集，具有有限的测度 $L(\Omega)>0$($L(\Omega)$ 表 Ω 的勒贝格测度，直观地说，对一维区间它是长度，对二维区域是面积，三维是体积 $\cdots$). 向 Ω 中投掷一质点 M, 如果 M 在 Ω 中均匀分布，那么就称这随机试验 (掷点) 是几何型的. 所谓 "M 在 Ω 中均匀分布" 的详细内容是："点 M 必定落于 Ω 中，而且落在可测集 $A(\subset\Omega)$ 中的可能性大小与 A 的测度成正比，而与 A 的位置及形状无关". 这时 Ω 中每一点 ω 是一基本事件："M 落于 ω 上", 故有无穷多个基本事件而不能用 (3). 如果以 $P(A)$ 表 "M 落在 A 中" 的概率，考虑到均匀分布性，自然应定义



$$P(A)=\frac{L(A)}{L(\Omega)},\tag{6}$$

其中 $L(A)$ 表 A 的测度，我们认为空集 $\emptyset$ 也是可测的而且 $L(\emptyset)=0$.

注意，这里概率 $P(A)$ 只是对可测集 A 才有定义，而不是对 Ω 的所有子集都有定义，因为有不可测的子集存在.



例 6 (约会问题) 二人约定于 0 到 T 时内在某地见面，先到者等 $t(t\leqslant T)$ 时后离去，试求二人能会见的概率 p.

解 以 x,y 分别表二人到达时刻，

$$0\leqslant x\leqslant T,\ 0\leqslant y\leqslant T,$$

这样的 (x,y) 构成边长为 T 的正方形 Ω. 二人能会见的充分条件是 $|x-y|\leqslant t$, 这条件决定 Ω 中一子集 A. 由 (6)

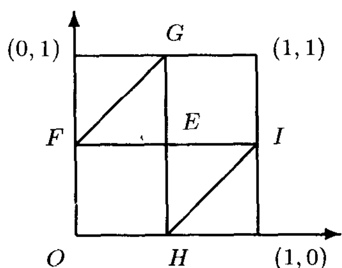
$$p=\frac{L(A)}{L(\Omega)}=\frac{T^2-(T-t)^2}{T^2}=1-\left(1-\frac{t}{T}\right)^2.\ \square$$

注意，例 6 可抽象化为：设有二点 A,B 分别几何型地落于 $[0,T]$ 中，试求 A,B 的距离不超过 $t(t\leqslant T)$ 的概率 p.

例 7 在线段 AD 上任意 [1] 取两点 B,C, 在 B,C 处折断此线段而得三折线，试求此三折线能构成三角形的概率.

解 设 AD 的长度为 l, A, D 的坐标各为 $(0,0),(l,0)$. 假定 B,C 的横坐标分别是 xl 及 yl, 那么 $0 \leqslant x,y \leqslant 1$. 由此可见 B,C 二点与正方形 $0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1$ 上的点 (x,y) 是一一对应的. 先设 $x < y$, 这时 AB, BC, CD, 构成三角形的充分必要条件是

$$\overline{AB}+\overline{BC}>\overline{CD},\quad \overline{BC}+\overline{CD}>\overline{AB},\quad \overline{CD}+\overline{AB}>\overline{BC},$$



其中 $\overline{AB}$ 表 AB 的长. 注意 $\overline{AB}=xl$, $\overline{BC}=(y-x)l$, $\overline{CD}=(1-y)l$, 代入上面的三个不等式中，得

$$y>\frac{1}{2},\quad x<\frac{1}{2},\quad y-x<\frac{1}{2}. \tag{7}$$

(7) 式决定正方形内一三角形 GEF, 它的面积是 $\frac{1}{8}$. 次设 $y<x$, 同样可得三角形 EHI, 面积也是 $\frac{1}{8}$. 注意 $x=y$ 时 B,C 重合而不可能得三角形. 这样一来，我们的问题等价于向边长为 1 的正方形上均匀分布地掷点而求点落于 ΔGFE 或 ΔEHI 中的概率. 由 (6) 得所求概率为 $2\times\frac{1}{8}=\frac{1}{4}$. 注意此概率与 l 无关. □

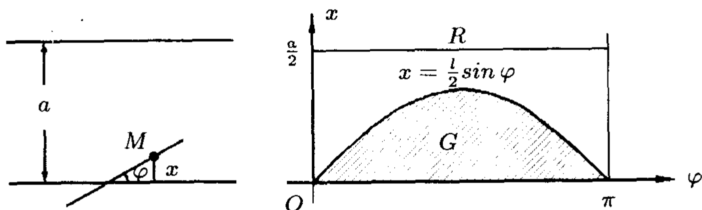
例 8 (蒲丰 [Buffon] 问题) 平面上画有等距离为 $a(a>0)$ 的一些平行线，向平面任意投一长为 $l(l<a)$ 的针，试求针与一平行线相交的概率 p.

解 以 M 表落下后针的中点， x 表 M 与最近一平行线的距离， φ 表针与此线的交角，易见

$$0 \leqslant x \leqslant \frac{a}{2},\quad 0 \leqslant \varphi \leqslant \pi.$$

这二式决定 $x\varphi$ 平面上一矩形 R; 其次，为了使针与一平行线 (这线必定是与 M 最近的平行线) 相交，充分与必要条件是

$$x \leqslant \frac{l}{2}\sin\varphi.$$



这不等式决定 R 中一子集 G. 因此，我们的问题等价于向 R 中均匀分布地掷点而求点落于 G 中的概率 p. 由 (6) 得

$$p=\int_0^\pi \frac{l}{2}\sin\varphi\, d\varphi \Big/ \frac{a}{2}\pi=\frac{2l}{\pi a}. \ \square \tag{8}$$

注意 p 只依赖于比值 $\frac{l}{a}$, 故当 l,a 成比例地变化时， p 的值不变，这正与直观符合. (8) 式提供了一个求 π 的值的方法：如果我们能事先求得 p, 那么由 (8) 就可以求出 π(续看 §2.9 例 7) [2].

---

[1] 所谓“任意”, 严格地说，应理解为“B 及 C 都在 AD 上独立地均匀分布”.

[2] 此实验已由多人做过，例如， 1850 年 Wolf 掷针 5000 次得 π 之估值为 3.1596; 1901 年 Lazzerini 掷 3408 次得估值为 3.1415929. 后者掷的次数少反而得到好的结果，原因是他在合适的次数上停止了试验. 但如事先不知 π 的真值时，应在何时停止实验？这是所谓“最佳停止问题”, 值得研究.

与定理 1 类似，可以证明下列定理：

定理 2 对几何型随机试验 E, 概率具有下列性质：

(i) 对任意可测集 A, $1 \geqslant P(A) \geqslant 0$;

(ii) $P(\Omega)=1$;

(iii) 设可列多个可测集 $A_1, A_2, \cdots$ 互不相交，则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i). \tag{9}$$

证 由定义中的 (6) 式立得 (i)(ii). 利用测度的完全可加性：对可列多个互不相交的可测集 $A_1, A_2, \cdots$, 有

$$L\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} L(A_i);$$

由此立得

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\frac{L\left(\bigcup\limits_{i=1}^{\infty} A_i\right)}{L(\Omega)}=\sum_i \frac{L(A_i)}{L(\Omega)}=\sum_i P(A_i). \ \square$$

(四) 频率. 设 E 为任一随机试验，A 为其中任一事件，在同样条件下，把 E 独立地重复做 n 次，以 $f_n(A)$ 表事件 A 在这 n 次试验中出现的次数，比值

$$F_n(A)=\frac{f_n(A)}{n}$$

称为事件 A 在这 n 次试验中出现的频率，而 $f_n(A)$ 称为 A 在这 n 次试验中出现的频数. 例如掷 100 次硬币中得到 51 次正面，那么“得正面”这一事件在这 100 次试验中的频率为 $\frac{51}{100}$, 频数为 51.

容易想到，一般地如 A 出现的可能性愈大，频率 $F_n(A)$ 也愈大，反之，如果 $F_n(A)$ 愈大，那么可以设想 A 出现的可能性也愈大. 因此，频率与概率间应有紧密的关系. 的确，以后可以证明：在相当广泛的条件下，当 $n \to \infty$ 时，在一定意义下 $F_n(A)$ 趋于 A 的概率 $P(A)$.

因此，当 n 充分大时，可以取频率作为概率的近似值. 在许多实际问题中，当概率不易求出时，往往就是这样做的.

频率的重大意义，这一方面是由于它能适当地反映 A 出现可能性的大小，另一方面，频率的概念比较简单，容易掌握，我们常常可以根据频率的性质去推想概率的性质. 这就是我们为什么在这里要引进频率的原因.

定理 3 对任意随机试验 E, 频率具有下列性质：

(i) 对任意事件 $A, 1 \geqslant F_n(A) \geqslant 0$;

(ii) $F_n(\Omega)=1$;

(iii) 对任意有穷多个互不相容的事件 $A_1, A_2, \cdots, A_m$, 即 $A_iA_j=\emptyset, i \neq j$, 有

$$F_n\left(\bigcup_{i=1}^{m} A_i\right)=\sum_{i=1}^{m} F_n(A_i), \tag{10}$$

其中 n 为任意正整数.

证 由 $F_n(A)$ 的定义 (i) 是显然的. 既然 Ω 是必然事件, $f_n(\Omega)=n$, 故得 (ii), 最后, 根据这些事件的互不相容性得

$$f_n\left(\bigcup_{i=1}^{m} A_i\right)=\sum_{i=1}^{m} f_n(A_i),$$

以 n 除上式两边就得 (10). □

作为频率的应用, 再来考虑蒲丰氏问题. 以 A 表事件: "针与一平行线相交". 如上所述, 当 n 充分大时, $F_n(A)$ 接近于 $P(A)$, 故由 (8) 得

$$\pi \approx \frac{2l}{F_n(A)a},$$

"$\approx$" 表 "近似". 因而我们可以用实验的方法来求 π 的近似值. 这种思想在计算数学的 Monte-Garlo 方法中得到广泛应用. (参看 §3.3 例 1 及第六, 七章)

关于几何概率的例, 还可见 §1.6.

定理 1-3 启示我们在一般情况下 (不限于古典型或几何型随机试验) 应该如何定义概率, 在 §1.3 中将讨论这个问题.

§1.2 古典型概率

(一) **复合随机试验**. 设已给 n 个随机试验 $E_1, E_2, \cdots, E_n$, 其中 E_i 的基本事件记为 $\omega^{(i)}$, 因而 $\Omega_i=(\omega^{(i)})$ 是 E_i 的基本事件空间. 我们可以把 $\{E_1, E_2, \cdots, E_n\}$ 看成为一个复合的随机试验 $\widetilde{E}$, 就是说, 把 $\widetilde{E}$ 做一次相当于把 $E_1, E_2, \cdots, E_n$ 顺次各做一次, 于是 $\widetilde{E}$ 的试验结果是由 n 次试验 $E_1, E_2, \cdots, E_n$ 的结果顺次联合组成的. 我们称 $\widetilde{E}$ 为由 $E_1, E_2, \cdots, E_n$ 组成的 n 次复合随机试验, 并记

$$\widetilde{E}=E_1 \times E_2 \times \cdots \times E_n=\prod_{i=1}^{n} E_i. \tag{1}$$

显然, $\widetilde{E}$ 的任一基本事件 $\widetilde{\omega}$ 是具有 n 个分量的点:

$$\widetilde{\omega}=(\omega^{(1)}, \omega^{(2)}, \cdots, \omega^{(n)}), \tag{2}$$

其中 $\omega^{(i)} \in \Omega_i$. 因此, $\widetilde{E}$ 的基本事件空间 $\widetilde{\Omega}$ 由一切这样的点所构成, 即

$$\widetilde{\Omega}=\{(\omega^{(1)}, \omega^{(2)}, \cdots, \omega^{(n)})\}, \tag{3}$$

有时记为

$$\widetilde{\Omega}=\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n=\prod_{i=1}^{n} \Omega_i. \tag{4}$$

特别, 当 $E_1, E_2, \cdots, E_n$ 重合于一个随机试验 E 时, 简记 $\widetilde{E}=E^n$, 并称 $\widetilde{E}$ 为 E 的 n 次重复随机试验, 这时 (2),(3) 及 (4) 分别化为

$$\widetilde{\omega}=(\omega_1, \omega_2, \cdots, \omega_n);\quad \widetilde{\Omega}=\{(\omega_1, \omega_2, \cdots, \omega_n)\}=\Omega^n, \tag{5}$$

其中 $\omega_i \in \Omega$.

例 1 设 E_1 为掷一枚硬币，$\Omega_1 = (\omega_1^{(1)}, \omega_2^{(1)})$, $\omega_1^{(1)}$ 表得正面，$\omega_2^{(1)}$ 表得反面；E_2 为自标号为 1,2,3 的三个球中任意抽取一球，$\Omega_2 = (\omega_1^{(2)}, \omega_2^{(2)}, \omega_3^{(2)})$, 其中 $\omega_i^{(2)}$ 表取得标号为 i 的球，$i = 1, 2, 3$. 于是对 2 次复合随机试验 $\widetilde{E} = E_1 \times E_2$, 做一次 $\widetilde{E}$ 相当于先掷一次硬币再抽取一球，故 $\widetilde{\Omega} = \Omega_1 \times \Omega_2$ 共含六个基本事件，$(\omega_1^{(1)}, \omega_1^{(2)})$, $(\omega_1^{(1)}, \omega_2^{(2)})$, $(\omega_1^{(1)}, \omega_3^{(2)})$, $(\omega_2^{(1)}, \omega_1^{(2)})$, $(\omega_2^{(1)}, \omega_2^{(2)})$, $(\omega_2^{(1)}, \omega_3^{(2)})$.

例 2 设 E 为掷一枚硬币，$\Omega = (\omega_1, \omega_2)$, $\omega_1(\omega_2)$ 表得正 (反) 面，将 E 重复 n 次而得 n 次重复随机试验 $\widetilde{E} = E^n$, 于是 $\widetilde{E}$ 的一个基本事件是由总共 n 个元："正面" 或 "反面", 构成的. 显然 Ω^n 共有 2^n 个不同的基本事件.

例 3 设 E_i 共有 m_i 个不同的基本事件，则 $\widetilde{E} = E_1 \times E_2 \times \cdots \times E_n$ 共有 $\prod\limits_{i=1}^{n} m_i$ 个不同的基本事件. 特别，如 $\widetilde{E} = E^n$, 且 E 共有 m 个不同的基本事件，那么 $\widetilde{E}$ 的不同的基本事件共有 m^n 个.

设已给可列多个随机试验 $E_1, E_2, \cdots$, 类似地可以考虑由它们组成的可列次复合随机试验

$$\widetilde{E} = E_1 \times E_2 \times \cdots = \prod_{i=1}^{\infty} E_i. \tag{6}$$

如果 $\Omega_i = (\omega^{(i)})$ 是 E_i 的基本事件空间，那么 $\widetilde{E}$ 的任一基本事件 $\widetilde{\omega}$ 是一点列：

$$\widetilde{\omega} = (\omega^{(1)}, \omega^{(2)}, \cdots), \ (\omega^{(i)} \in \Omega_i). \tag{7}$$

因而 $\widetilde{E}$ 的基本事件空间 $\widetilde{\Omega}$ 是所有可能的点列的集

$$\widetilde{\Omega} = \{(\omega^{(1)}, \omega^{(2)}, \cdots)\} \tag{8}$$

有时记为

$$\widetilde{\Omega} = \Omega_1 \times \Omega_2 \times \cdots = \prod_{i=1}^{\infty} \Omega_i. \tag{9}$$

特别，当 $E_i = E$ 时，$(i = 1, 2, \cdots)$, 简记 $\widetilde{E} = E^\infty$, 并称 $\widetilde{E}$ 为 E 的可列次重复随机试验，这时 (7) 及 (8)(9) 分别化为

$$\widetilde{\omega} = (\omega_1, \omega_2, \cdots); \ \ \widetilde{\Omega} = \{(\omega_1, \omega_2, \cdots)\} = \Omega^\infty, \ (\omega_i \in \Omega). \tag{10}$$

例如，考虑例 2 中的 E, 这时 E^∞ 中任一点是由一列 "ω_1- 正面""ω_2- 反面" 构成的. 譬如说

$$(\omega_1, \omega_1, \omega_2, \omega_1, \omega_2, \cdots)$$

就是一基本事件，而 Ω^∞ 是所有这种序列的集.

(二) **排列与组合**. 为了计算古典型随机试验中事件的概率，需要一些排列与组合的基本知识.

三个数字 1,2,3 共有 6 种不同的排列：123,132,213,231,312,321. 一般地，试问 n 个不同的元共有几种排列？每个排列共有 n 个位置，第一个位置上可以放进这 n 个元中的任何

一个，共有 n 种放法；第二个位置上可以放进剩下的 $n-1$ 个元中的任何一个，共有 $n-1$ 种放法；…….因此，共有

$$n! = n(n-1)(n-2)\cdots 2\cdot 1 \tag{11}$$

种不同的排列.

更一般地，从 n 个不同元 $a_1, a_2, \cdots, a_n$ 中任取 $m(m \leqslant n)$ 个，并把它们排列，试问共有几种排列？利用类似的想法，第一个位置上可以有 n 种放法，第二个位置上有 $n-1$ 种放法，…，第 m 个位置上有 $n-m+1$ 种放法，因此，共有

$$n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!} \tag{12}$$

种不同的排列.

在排列中，我们不仅注意了参加排列的是哪些元，而且还注意了元的次序，例如 $a_1a_2a_3$ 与 $a_1a_3a_2$ 虽然有相同的元，但因有不同的次序而被看成为两个不同的排列.

如果从 n 个不同的元 $a_1, a_2, \cdots, a_n$ 中任取 $m(m \leqslant n)$ 个而不论它们的次序，我们就得到了一种组合；两个组合只要有相同的元，不论次序如何，都认为是一样的，例如 $a_1, a_2, \cdots, a_{m-1}a_m$ 与 $a_m, a_{m-1}, \cdots, a_2a_1$ 是同一组合. 由于一种组合对应于 $m!$ 种排列，所以共有

$$\frac{n(n-1)\cdots(n-m+1)}{m!} = \frac{n!}{m!(n-m)!} \tag{13}$$

种不同的组合. 可以把组合问题换一种说法：要把 n 个不同的元 $a_1, a_2, \cdots, a_n$ 分成两组，一组 $m(m \leqslant n)$ 个，另一组 $n-m$ 个，那么不同的分法共有 $\frac{n!}{m!(n-m)!}$ 种. 通常[1]把 $\frac{n!}{m!(n-m)!}$ 记成 $\binom{n}{m}$.

一般地，要把 n 个不同的元分成 $k(1 \leqslant k \leqslant n)$ 组，使第 i 组含 m_i 个元，因而有 $\sum\limits_{i=1}^{k} m_i = n$, 那么，可以先把 n 个分成两组，甲组含 m_1 个，乙组含 $n-m_1$ 个，再把乙组分成两小组，其一含 m_2 个，另一含 $n-m_1-m_2$ 个，……，因而总共不同的分法种数是

$$\frac{n!}{m_1!(n-m_1)!}\frac{(n-m_1)!}{m_2!(n-m_1-m_2)!}\cdots\frac{(n-m_1-\cdots-m_{k-2})!}{m_{k-1}!(n-m_1-\cdots-m_{k-1})!} = \frac{n!}{m_1!m_2!\cdots m_k!}. \tag{14}$$

这时，如果两个分法中，第 1 组 m_1 个元彼此一样，…，第 k 组 m_k 个元也彼此一样，那么便认为两个分法相同.

(三) 古典型随机试验中的概率计算. 现在举一些例子来说明如何计算古典型中事件 A 的概率. 实际中许多具体问题可以大致归并为三类，它们具有典型的意义：

(I) 抽球问题；(II) 分房问题；(III) 随机取数问题.

以下的例子如果属于 (i) 类，我们就标以 (i), i =I,II,III.

计算的步骤是：首先应求出不同的基本事件的个数 n, 因而要弄清基本事件空间是什么，其次须求出所研究的事件 A 所含不同的基本事件的个数 k, 于是 $P(A)=\frac{k}{n}$.

---

[1]为方便计，如 $m > n$, 定义 $\binom{n}{m} = 0$.

例 4 (I) 箱中盛有 α 个白球及 β 个黑球，从其中任意取 $a+b$ 个，试求所取的球恰含 a 个白球和 b 个黑球的概率.

解 这里的随机试验 E 是自 $\alpha+\beta$ 个球中取出 $a+b$ 个，每 $a+b$ 个球构成一基本事件 ω, 所以共有 $\binom{\alpha+\beta}{a+b}$ 个不同的 ω. 事件 A:"恰好取中 a 个白球 b 个黑球" 共含 $\binom{\alpha}{a}\times\binom{\beta}{b}$ 个不同的 ω, 故

$$P(A)=\frac{\binom{\alpha}{a}\cdot\binom{\beta}{b}}{\binom{\alpha+\beta}{a+b}}. \quad \square \tag{15}$$

例 5 (I) 箱中盛有 α 个白球和 β 个黑球，从其中任意地接连取出 $k+1(k+1\leqslant\alpha+\beta)$ 个球，如每球被取出后不还原，试求最后取出的球是白球的概率.

解 这里 E 是自 $\alpha+\beta$ 个球中接连不还原地取出 $k+1$ 个球，由于注意了球的次序，故应考虑排列. 每 $k+1$ 个排列好的球构成一基本事件，接连不还原取 $k+1$ 个可看成一次取 $k+1$ 个，所以共有 $\binom{\alpha+\beta}{k+1}$ 种不同取法，一种取法对应 $(k+1)!$ 个排列，故共有 $\binom{\alpha+\beta}{k+1}\cdot(k+1)!$ 个不同的 ω, 所需白球可以是 α 个中的任一个，有 α 种选法，其余 k 个可以是其余 $\alpha+\beta-1$ 个中的任意 k 个，有 $\binom{\alpha+\beta-1}{k}$ 种选法. 因而事件 A:"取出的 $k+1$ 个球中最后一球是白球" 共含 $\alpha\binom{\alpha+\beta-1}{k}\cdot k!$ 个不同的基本事件，于是

$$P(A)=\frac{\alpha\binom{\alpha+\beta-1}{k}\cdot k!}{\binom{\alpha+\beta}{k+1}\cdot(k+1)!}=\frac{\alpha}{\alpha+\beta}. \tag{16}$$

值得注意的是 $P(A)$ 与 k 无关. $\square$

以后还会遇到各种各样的抽球问题，注意，这里的"白球"、"黑球"可换为"甲物"、"乙物"或"合格产品","不合格产品"等等，我们所以说抽球问题有典型意义，原因就在于此.

例 6 (II) 有 n 个人，每个人都以同样的概率 $\frac{1}{N}$ 被分配在 $N(n\leqslant N)$ 间房中的每一间中，试求下列各事件的概率：

A: 某指定 n 间房中各有一人；

B: 恰有 n 间房，其中各有一人；

C: 某指定房中恰有 $m(m\leqslant n)$ 人.

解 设 E 为随机试验："把一人分配到 N 间房中之一去", 于是每间房对应于一基本事件而 E 共有 N 个不同的基本事件. 本例中的随机试验 $\widetilde{E}=E^n$, 故由第 (一) 段知 $\widetilde{E}$ 有 N^n 个不同的基本事件 $\widetilde{\omega}$.

今固定某 n 间房，第一人可分配到其中任一间，故有 n 种分法，第二人可分配到余下 $n-1$ 间中任一间，故有 $n-1$ 种分法，$\cdots\cdots$ 因而事件 A 共含 $n!$ 个不同的 $\widetilde{\omega}$. 于是

$$P(A)=\frac{n!}{N^n}. \tag{17}$$

如果这 n 间房可自 N 间中任意选出，那么共有 $\binom{N}{n}$ 种选法，因而事件 B 共含 $\binom{N}{n}n!$ 个不同的 $\widetilde{\omega}$, 于是

$$P(B)=\frac{\binom{N}{n}n!}{N^n}=\frac{N!}{N^n(N-n)!}. \tag{18}$$

事件 C 中的 m 个人可自 n 个人中任意选出，共有 $\binom{n}{m}$ 种选法，其余 $n-m$ 个人可以任意分配在其余的 $N-1$ 间房里，共有 $(N-1)^{n-m}$ 个分配法，因而事件 C 共有 $\binom{n}{m}\cdot(N-1)^{n-m}$ 个不同的 $\tilde{\omega}$, 于是

$$P(C)=\frac{\binom{n}{m}(N-1)^{n-m}}{N^n}=\binom{n}{m}\left(\frac{1}{N}\right)^m\left(\frac{N-1}{N}\right)^{n-m}. \tag{19}$$

这是所谓二项分布的特殊情形 (见 §2.3), 注意当 n 及 N 固定时， $P(C)$ 只依赖于 m, 如果记 $P(C)$ 为 P_m, 则

$$\sum_{m=0}^{n}p_m=\sum_{m=0}^{n}\binom{n}{m}\left(\frac{1}{N}\right)^m\left(\frac{N-1}{N}\right)^{n-m}=\left(\frac{1}{N}+\frac{N-1}{N}\right)^n=1. \ \square$$

许多表面上提法不同的问题实际上属于同一类型，试看下例：

例 7 (II) (i) 有 n 个质点，设每个都以同样的概率 $\frac{1}{365}$ 落于 365 个格子中的每一格中，试求每一格至多只含一点的概率 p.

(ii) 有 n 个人，设每人的生日是任何一日的概率为 $\frac{1}{365}$, 试求此 n 人的生日互不相同的概率 p'.

(iii) 有 n 个旅客乘火车途经 365 站，设每人在每站下车的概率为 $\frac{1}{365}$, 试求没有一人以上同时下车的概率 p''.

在 (i)(ii)(iii) 中都假定 $n\leqslant 365$.

解 这三个问题其实都和例 6 中求 $P(B)$ 的问题等价 (取那里的 $N=365$), 只要把 “人” ,“质点”,“旅客” 看成一样，把 “房”,“格子”,“日”,“站” 看成一样，因而由例 6 的解答知

$$p=p'=p''=P(B)=\frac{365\cdot 364\cdots(365-n+1)}{365^n}. \tag{20}$$

例如当 $n=30$ 时， $P(B)=0.294$, 而 B 的对立事件 $\overline{B}$ 的概率由 §1.1(5) 为

$$P(\overline{B})=1-P(B)=0.706,$$

这里三位以后的小数略去. 由此可见， 30 人中，至少有二人同生日的概率大于 70%.

要准确地算出 $P(B)$, 计算量一般是很大的，试求近似值如下：改写 (20) 为

$$P(B)=\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{n-1}{365}\right). \tag{21}$$

当 n 较小时，把含有一个以上因子的乘积如 $\frac{1}{365}\cdot\frac{2}{365}$ 等项略去，得

$$P(B)\approx 1-\frac{1+2+\cdots+(n-1)}{365}=1-\frac{n(n-1)}{730}, \tag{22}$$

这里 “≈” 表 “近似于”. 另一近似公式如下：注意对小的正数 x 有 $\log(1-x)\approx -x$, 故由 (21) 得

$$\log P(B)\approx -\frac{1+2+\cdots+(n-1)}{365}=-\frac{n(n-1)}{730}. \tag{23}$$

例如，当 $n=10$, (21) 式给出 $P(B)=0.883\cdots$, 而 (22) 给出 $P(B)\approx 0.877$. $\square$

例 8 (III) 从 $1,2,\cdots,10$ 共十个数字中任取一个，假定每个数字都以 $\frac{1}{10}$ 的概率被取中，取后还原，先后取出 7 个数字，试求下列各事件 $A_i\ (i=1,2,\cdots,5)$ 的概率：

A_1: 7 个数字全不相同；

A_2: 不含 10 与 1;

A_3: 10 恰好出现两次；

A_4: 至少出现两次 10;

A_5: 总和为 20.

解 设取一次所成的随机试验为 E, 它有 10 个不同的基本事件，因而还原地取 7 次所成的随机试验是 7 次重复的，$\widetilde{E}=E^7$. $\widetilde{E}$ 共有 10^7 个不同的基本事件 $\widetilde{\omega}$. 容易看出

$$P(A_1)=\frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4}{10^7}\approx 0.06048,$$

$$P(A_2)=\frac{8^7}{10^7}.$$

今计算 $P(A_3)$. 出现 10 的两次可以是 7 次中的任意二次，故有 $\binom{7}{2}$ 种选择，其他 5 次中，每次只能取 9 个剩下数字中的任何一个，故

$$P(A_3)=\frac{\binom{7}{2}\cdot 9^5}{10^7}.$$

一般地， 10 恰好出现 $k(k\leqslant 7)$ 次的概率 p_k 为

$$p_k=\frac{\binom{7}{k}\cdot 9^{7-k}}{10^7}. \tag{24}$$

由于 A_4 是 6 个互不相容的事件 “10 恰好出现 k 次”$(k=2,3,4,5,6,7)$ 的和，故由 (24) 及 §1.1(5) 得

$$P(A_4)=\sum_{k=2}^{7}p_k=\frac{\sum_{k=2}^{7}\binom{7}{k}\cdot 9^{7-k}}{10^7}.$$

现在计算 $P(A_5)$.A_5 所含的不同的 $\widetilde{\omega}$ 的个数是下列方程的整数解的个数.

$$\sum_{i=1}^{7}x_i=20\quad(1\leqslant x_i\leqslant 10), \tag{25}$$

其中 x_i 代表第 i 次所取得的数字. 然而 (25) 的上述解的个数重合于 $(x+x^2+x^3+\cdots+x^{10})^7$ 中 x^{20} 的系数，亦即 $(1+x+x^2+\cdots+x^9)^7$ 中 x^{13} 的系数. 由于

$$\begin{aligned}(1+x+x^2+\cdots+x^9)^7&=\left(\frac{1-x^{10}}{1-x}\right)^7=(1-x^{10})^7(1-x)^{-7}\\&=(1-7x^{10}+\cdots)(1+\cdots+84x^3+\cdots+27132x^{13}+\cdots),\end{aligned}$$

可见 x^{13} 的系数为 $27132-7\times 84=26544$. 所以

$$P(A_5)=\frac{26544}{10^7}.\ \square$$

最后叙述一个在统计物理中起重要作用的例，这例还说明，在具体问题中，如何计算基本事件的个数需根据实际情况而定，因而赋予概率的方法也应如此.

例 9 (II) 设有 n 个质点，每点都以概率 $\frac{1}{N}$ 落于 $N(\geqslant n)$ 个盒子中的每一个里. 试求事件 A:"某预先指定的 n 个盒中各含一点" 的概率 p.

由于对质点和盒子所作的进一步的假定不同，这题有三种不同的解法和答案. 相对于这些假定来说，每种解法都是正确的.

解 1 (Maxwell-Boltzmann) 假定 n 个质点是不同的，即可辨别的；还假定每个盒子能容纳的质点数是没有限制的. 根据例 6 得

$$p=\frac{n!}{N^n}.$$

解 2 (Bose-Einstein) 假定 n 个质点完全相同，因而不可辨别；对每个盒子则仍然假定它能容纳任意多个质点. 在解 1 中，分布法不仅依赖于每个盒子中的质点的个数，而且还依赖于是那几个质点；而现在则只依赖于盒中的质点个数，因此计算基本事件个数的方法不同了. 为了更清楚地说明这点，我们就 $N=3, n=2$ 的情形来写出这二种解法中的全体基本事件. 以 a, b 表这二个质点，下面的记号 $(ab,-,-)$ 表示第一盒中含二点，其余皆空. 其它记号的解释类似，于是解 1 中共有下列 $3^2=9$ 个基本事件：

1 $(ab,-,-)$	2 $(-,ab,-)$	3 $(-,-,ab)$
4 $(a,b,-)$	5 $(b,a,-)$	6 $(a,-,b)$
7 $(b,-,a)$	8 $(-,a,b)$	9 $(-,b,a)$

但在解 2 的假定中，则不区别 4,5 而把它们看成为同一基本事件，因为 $a=b$. 同样，6,7 与 8,9 也如此. 故共只有 6 个基本事件，而 A 则只含一个 (如预先指定的是第一，二个盒子，则 A 所含的基本事件由 4,5 表示)，于是在 $N=3, n=2$ 时，$p=\frac{1}{6}$.

回到一般情况，我们来计算基本事件的个数. 为此，采用下列巧妙的直观想法. 把 N 个盒子并排成一行. n 个点的一种分布法可表为下图：

$$|*\ *|*|*\ *\ *|-|\cdots|*\ *|$$

这表示第一盒有 2 点，第二盒 1 点，第三盒 3 点，第四盒空，最后一盒 2 点. $*$ 表点，"|" 表盒子的壁. 当然，最外两端各有一壁. 把每个壁与每个点都看成一个位置，则 n 个球的一种分布法就相当于 $n+N-1$ 个位置 (两端的不在内) 被 n 个点占领的一种占位法，故共有 $\binom{n+N-1}{n}$ 种分布法，亦即共有 $\binom{n+N-1}{n}$ 个基本事件. 事件 A 只含其中的一个，故

$$p=\frac{1}{\binom{n+N-1}{n}}=\frac{n!(N-1)!}{(n+N-1)!}.$$

实际中的确有许多质点被认为是不可辨别的，例如微观世界中同类的基本粒子 (如电子或光子)，至少在目前的科学水平上就是如此，因为现今还不能把两个电子区别开来. 关于这问题的进一步讨论见第一章习题 14 及其解答. 解 2 是物理学家 Bose 与 Einstein 于 1924 年提出的，它适用于所谓玻色子的基本粒子.

解 3 (Fermi-Dirac) 这时假定每个盒子至多只能容纳一个质点；而对质点则仍设为不可辨别的. 于是任一种分布法都必须占用 n 个盒子. 由于质点不可辩，分布法只依赖于

各盒中的落点个数，而不依赖于落的是那几个点. 因而共有 $\binom{N}{n}$ 种分布法，亦即共有 $\binom{N}{n}$ 个基本事件. 事件 A 只含其中的一个，故

$$p = \frac{1}{\binom{N}{n}} = \frac{n!(N-n)!}{N!}.$$

解 3 是物理学家 Fermi 与 Dirac 于 1925 年提出的，它适用于所谓费米子的基本粒子.

上述各例表明：在计算古典型概率时，关键在于如何根据问题的条件区分两个不同的基本事件；只有这样才能正确计算出该试验中所有不同的基本事件的总数 n, 以及所考虑的事件 A 中所含不同的基本事件的个数 k, 二者的比 $\frac{k}{n}$ 就是所求的概率 p.

§1.3 概 率 空 间

(一) **概率的公理化定义**. 到现在为止，我们只对古典型及几何型的随机试验定义了概率，然而这两种随机试验远远没有穷尽所有的随机试验，譬如说，§1.1 中例 3 与例 4 就不属于这两种类型，于是产生在一般情况下概率应如何定义的问题.

我们已看到，上述两种类型中的概率具有公共的性质 (即 §1.1 定理 1 及 2), 特别重要的是：任一随机试验中的的频率也有类似的性质，(§1.1 定理 3). 这使人想到：可以用这些性质来作为一般的概率的定义. 近代概率论的公理结构中正是这样做的. 由于这些性质中只涉及空间 Ω 及其中的事件，所以在抽象化的场合，我们甚至可以摆脱具体的随机试验而直接从事件的概念开始. 现在就来叙述概率论的公理结构，它给出了事件与概率的严格定义.

定义 1 设 Ω 是抽象的点 ω 的集，$\Omega = (\omega)$, Ω 中的 **一些** 子集[1] A 所成的集 $\mathcal{F}$ 称为 Ω 中的一个 $\sigma-$ 代数，如果 $\mathcal{F}$ 满足下列条件：

(i) $\Omega \in \mathcal{F}$;

(ii) 如 $A \in \mathcal{F}$, 则 $\overline{A} \in \mathcal{F}$, $(\overline{A} = \Omega - A)$;

(iii) 如可列多个 $A_m \in \mathcal{F}$, $m = 1, 2, \cdots$, 则

$$\bigcup_{m=1}^{\infty} A_m \in \mathcal{F}.$$

定义 2 设 $P(A)(A \in \mathcal{F})$ 是定义在 $\sigma-$ 代数 $\mathcal{F}$ 上的实值集函数，如果它满足下列条件，就称它为 $\mathcal{F}$ 上的概率测度，或简称概率.

1° 对每 $A \in \mathcal{F}$, 有 $1 \geqslant P(A) \geqslant 0$;

2° $P(\Omega) = 1$;

3° 如可列多个 $A_m \in \mathcal{F}$, $A_i A_j = \emptyset, i \neq j$, 则

$$P\left(\bigcup_{m=1}^{\infty} A_m\right) = \sum_{m=1}^{\infty} P(A_m). \tag{1}$$

称点 ω 为基本事件，$\mathcal{F}$ 中的集 A 称为事件，因而 $\mathcal{F}$ 是全体事件的集，$P(A)$ 称为事件 A 的概率，三元总体 $(\Omega, \mathcal{F}, P)$ 称为概率空间.

[1]空集 $\emptyset$ 是任一集的子集，Ω 也是 Ω 的子集.

例 1 设 $\Omega=(\omega_1,\omega_2,\cdots,\omega_n)$ 只含 n 个不同的基本事件，Ω 的全体子集的集记为 $\mathcal{F}$, 因而 $\mathcal{F}=\{\emptyset;(\omega_1),(\omega_2),\cdots,(\omega_n);(\omega_1,\omega_2),(\omega_1,\omega_3),\cdots,(\omega_{n-1},\omega_n),,\cdots,\Omega\}$ 共含 $\sum\limits_{i=0}^{n}\binom{n}{i}=(1+1)^n=2^n$ 个不同的集，($\binom{n}{0}=1$). 由于 $\mathcal{F}$ 包含了 Ω 的所有子集，显然条件 (i)(ii)(iii) 满足，故此 $\mathcal{F}$ 是 Ω 中的一个 $\sigma-$ 代数. 对任意 $A\in\mathcal{F}$, 定义

$$P(A)=\frac{k}{n}, \tag{2}$$

其中 k 是 A 中不同的基本事件的个数，根据 §1.1 定理 1, 可见条件 1°, 2°, 3° 满足 (注意这时 Ω 中不可能有可列多个非空的互不相交的集), 因而 $(\Omega,\mathcal{F},P)$ 是一概率空间，这正是古典型随机试验的概率空间.

例 2 设 $\Omega=(\omega)$ 是 n 维空间中具有有限，正勒贝格 (Lebesgue) 测度的集，Ω 中一切 (勒贝格) 可测子集构成一集 $\mathcal{F}$, $\mathcal{F}$ 是满足 (i),(ii),(iii) 的，对 $A\in\mathcal{F}$, 定义

$$P(A)=\frac{L(A)}{L(\Omega)}, \tag{3}$$

这里 $L(A)$ 表 A 的勒贝格测度. 由 §1.1 定理 2, 可见 1°, 2°, 3° 满足，因而 $(\Omega,\mathcal{F},P)$ 是一概率空间，这正是几何型随机试验的概率空间.

例 3 设 $\Omega=(0,1,2,\cdots)$ 由全体非负整数构成，Ω 的全体子集所成的集 $\mathcal{F}$ 是 Ω 中的一个 $\sigma-$ 代数，对任意 $A\in\mathcal{F}$, 定义

$$P(A)=\sum_{k\in A}e^{-\lambda}\frac{\lambda^k}{k!},\quad (\lambda>0 \text{ 常数}) \tag{4}$$

$$P(\emptyset)=0. \tag{5}$$

我们来验证 1°, 2°, 3° 满足：实际上，1° 是显然的；

$$P(\Omega)=\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot e^{\lambda}=1; \tag{6}$$

如 $A_m\in\mathcal{F}$, $m=1,2,\cdots$, $A_iA_j=\emptyset$, 由 (4)(5) 有

$$P\left(\bigcup_{m=1}^{\infty}A_m\right)=\sum_{k\in\bigcup\limits_m A_m}e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{m=1}^{\infty}\sum_{k\in A_m}e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{m=1}^{\infty}P(A_m).$$

由此可见，$(\Omega,\mathcal{F},P)$ 是一概率空间.

由 $\sigma-$ 代数 $\mathcal{F}$ 的定义，可见任一概率空间具有下列性质：

(A) $\emptyset\in\mathcal{F}$.

事实上，在 (ii) 中取 $A=\Omega$, 立知 $\emptyset=\overline{\Omega}\in\mathcal{F}$.

(B) 如可列多个 $A_m\in\mathcal{F}$, $m=1,2,\cdots$, 则

$$\bigcap_{m=1}^{\infty}A_m\in\mathcal{F}.$$

事实上，如能证

$$\bigcap_{m=1}^{\infty} A_m = \overline{\bigcup_{m=1}^{\infty} \overline{A}_m}, \tag{7}$$

则由 (ii) 及 (iii) 即得知 $\bigcap\limits_{m=1}^{\infty} A_m \in \mathcal{F}$. (7) 式等价于下式

$$\overline{\bigcap_{m=1}^{\infty} A_m} = \bigcup_{m=1}^{\infty} \overline{A}_m. \tag{7'}$$

如 $\omega \in \overline{\bigcap\limits_{m=1}^{\infty} A_m}$, 也说是说，$\omega$ 不属于 $\bigcap\limits_{m=1}^{\infty} A_m$, 因而至少存在一 k, 使 ω 不属于 A_k, 亦即 $\omega \in \overline{A}_k$, 故 $\omega \in \bigcup\limits_{m=1}^{\infty} \overline{A}_m$, 这得证 $\overline{\bigcap\limits_{m=1}^{\infty} A_m} \subset \bigcup\limits_{m=1}^{\infty} \overline{A}_m$. 反之，将上述各步逆转，就得 $\overline{\bigcap\limits_{m=1}^{\infty} A_m} \supset \bigcup\limits_{m=1}^{\infty} \overline{A}_m$. 于是式 (7) 得证.

(7') 式是事件 (或集合) 运算中的重要关系，值得注意. 它表明："交的补集，等于补集的和", 而 (7) 式则表明："和的补集，等于补集之交". 这两句话是等价的.

(C) 如有穷多个 $A_m \in \mathcal{F}, m = 1, 2, \cdots, n$, 则 $\bigcup\limits_{m=1}^{n} A_m \in \mathcal{F}$, $\bigcap\limits_{m=1}^{n} A_m \in \mathcal{F}$.

事实上，在 (iii) 中取 $A_{n+1} = A_{n+2} = \cdots = \emptyset$, 即得 $\bigcup\limits_{m=1}^{n} A_m \in \mathcal{F}$, 在 (B) 中取 $A_{n+1} = A_{n+2} = \cdots = \Omega$, 即得 $\bigcap\limits_{m=1}^{n} A_m \in \mathcal{F}$,

(D) 如 $A \in \mathcal{F}, B \in \mathcal{F}$, 则 $A - B = A\overline{B} \in \mathcal{F}$.

关于概率空间的定义，我们再作一点注解. Ω 中的一集 A 是否一事件，完全决定于 A 是否属于 $\mathcal{F}$. 在定义 $\mathcal{F}$ 时，并没有要求 Ω 的全体子集都属于 $\mathcal{F}$, 因而并不是任何一集 A 都一定是一事件，我们只要求 $\mathcal{F}$ 满足 (i),(ii),(iii) 就行了. 这带来了很大的广泛性，否则连几何型中所定义的概率都不能纳入一般的概率定义中，因为 (3) 中的 $P(A)$ 只是对可测集 A 才有定义，而可测集不能穷尽 Ω 的全体子集. 那么，在实际问题中，应该如何选择 $\mathcal{F}$? 也就是说，应该挑选 Ω 的那些子集作为事件？这决定于问题的特殊性，必须具体问题具体解决. 一般如 Ω 只含有穷或可列多个点时，总是把 Ω 的全体子集的集取作 $\mathcal{F}$, 像例 1 和例 3 那样；如 Ω 不是有穷集或可列集时，就应根据问题的性质而选定 $\mathcal{F}$, 像例 2 那样. 当然，在有些问题里也可能把 $\mathcal{F}$ 取作全体子集的集. 好在以后在一般的理论中，我们总是事先假设 $\mathcal{F}$ 已经选定，因此这问题不会发生.

(二) 概率的性质.

定理 1 设 P 为概率，则

1. $P(\emptyset) = 0$;
2. 如 $A_m \in \mathcal{F}, m = 1, 2, \cdots, n$ 又 $A_iA_j = \emptyset, i \neq j$, 则

$$P\left(\bigcup_{m=1}^{n} A_m\right) = \sum_{m=1}^{n} P(A_m). \tag{8}$$

3. 对任意二事件 A_1, A_2, 有

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2). \tag{9}$$

证 1. 因 $\emptyset=\emptyset\cup\emptyset\cup\cdots$, 由 (1) 得 $P(\emptyset)=P(\emptyset)+P(\emptyset)+\cdots$, 于是 $P(\emptyset)=0$.

2. 在 3° 中取 $A_{n+1}=A_{n+2}=\cdots=\emptyset$. 利用 $P(\emptyset)=0$, 得

$$P\left(\bigcup_{m=1}^{n}A_m\right)=P\left(\bigcup_{m=1}^{\infty}A_m\right)=\sum_{m=1}^{\infty}P(A_m)=\sum_{m=1}^{n}P(A_m).$$

3. 因 $A_1\cup A_2=A_1\cup A_2\overline{A}_1, A_1\cap(A_2\overline{A}_1)=\emptyset$, 故

$$P(A_1\cup A_2)=P(A_1)+P(A_2\overline{A}_1). \tag{10}$$

然而， $A_2=A_1A_2\cup A_2\overline{A}_1, (A_1A_2)\cap(A_2\overline{A}_1)=\emptyset$, 故

$$P(A_2)=P(A_1A_2)+P(A_2\overline{A}_1).$$

由上式及 (10) 即得 (9) . □

系 1. 对任意 n 个事件 $A_1,A_2,\cdots,A_n$, 有

$$P\left(\bigcup_{m=1}^{n}A_m\right)\leqslant\sum_{m=1}^{n}P(A_m). \tag{11}$$

2. 如 A,B 为二事件， $A\supset B$, 则

$$P(A-B)=P(A)-P(B),\quad P(A)\geqslant P(B). \tag{12}$$

3. 对任意事件 A, 有

$$P(\overline{A})=1-P(A). \tag{13}$$

证 1. 当 $n=2$ 时，由 (9) 并注意 $P(A_1A_2)\geqslant 0$ 即得 (11); 设 $n=k$ 时 (11) 成立，令 $\bigcup_{m=1}^{k}A_m=C, A_{k+1}=D$, 利用刚才对 $n=2$ 证明的结果及 $n=k$ 时的假定得

$$\begin{aligned}P(C\cup D)&\leqslant P(C)+P(D)=P\left(\bigcup_{m=1}^{k}A_m\right)+P(A_{k+1})\\&\leqslant\sum_{k=1}^{k}P(A_m)+P(A_{k+1})=\sum_{m=1}^{k+1}P(A_m).\end{aligned}$$

2. 由 $A\supset B$ 得 $A=B\cup(A-B)$, 其中 $B\cap(A-B)=\emptyset$, 故 $P(A)=P(B)+P(A-B)$. 于是 (12) 中前式成立；后式则由前式及 $P(A-B)\geqslant 0$ 得到.

3. 因为 $\Omega=A\cup\overline{A}, A\overline{A}=\emptyset$, 所以 $1=P(\Omega)=P(A)+P(\overline{A})$, 从而得证 (13) . □

*(9) 式的一般化如下：对任意 n 个事件 $A_1,\cdots,A_n$, 令

$$\begin{aligned}&s_1=\sum_{i}P(A_i),\quad s_2=\sum_{i,j}P(A_iA_j),\\&s_3=\sum_{i,j,k}P(A_iA_jA_k),\ \cdots,\quad s_n=P(A_1A_2\cdots A_n),\end{aligned}$$

其中求和号对一切满足 $1 \leqslant i < j < k < \cdots \leqslant n$ 的 $i,j,k,\cdots$ 进行，例如 s_3 中的 $\sum\limits_{i,j,k}$ 表对一切满足 $1 \leqslant i < j < k \leqslant n$ 的正整数 i,j,k 求和，因而 i,j,k 各不相等，$P(A_1A_2A_3), P(A_1A_3A_2), \cdots$ 等六项只在此和中出现一次，即只论组合数. 故 s_r 共含 $\binom{n}{r}$ 项.

***系 2 (一般加法公式)**. 对任意 n 个事件 $A_1, A_2, \cdots, A_n$, 有

$$P\left(\bigcup_{i=1}^{n} A_i\right) = s_1 - s_2 + s_3 - s_4 + \cdots \pm s_n. \tag{14}$$

证 用归纳法，当 $n=2$ 时，(14) 化为 (9). 今设 (14) 对 $n=k$ 正确，由此及 (9) 得

$$P\left(\bigcup_{i=1}^{k+1} A_i\right) = P\left(\left[\bigcup_{i=1}^{k} A_i\right] \cup A_{k+1}\right) = P\left(\bigcup_{i=1}^{k} A_i\right) + P(A_{k+1}) - P\left(\bigcup_{i=1}^{k} A_i A_{k+1}\right)$$

$$= s_1 - s_2 + s_3 - s_4 + \cdots \pm s_k + P(A_{k+1}) - P\left(\bigcup_{i=1}^{k} B_i\right),$$

其中 $B_i = A_i A_{k+1}$, 对 $P\left(\bigcup\limits_{i=1}^{k} B_i\right)$ 用 (14), 对应地有

$$P\left(\bigcup_{i=1}^{k} B_i\right) = s_1' - s_2' + s_3' - s_4' + \cdots \pm s_k',$$

这里 $s_1' = \sum\limits_{i} P(A_i A_{k+1}), \cdots$, 以此代入上式，将 $P(A_{k+1})$ 与 s_1 合并，将 s_1' 与 s_2 合并，s_2' 与 s_3 合并，$\cdots$, 即知 (14) 对 $n=k+1$ 正确. □

定理 2 (连续性定理) 设 $A_n \in \mathcal{F}$, $A_n \supset A_{n+1}$, $n=1,2,\cdots$, 令 $A = \bigcap\limits_{n=1}^{\infty} A_n$, 则

$$P(A) = \lim_{n\to\infty} P(A_n). \tag{15}$$

证 由 $\{A_n\}$ 的单调不增性 $A_n \supset A_{n+1}$, 对任一 n 有

$$A_n = \left(\bigcup_{m=n}^{\infty} A_m \overline{A}_{m+1}\right) \cup \bigcap_{m=n}^{\infty} A_m = \left(\bigcup_{m=n}^{\infty} A_m \overline{A}_{m+1}\right) \cup A.$$

上式右方诸集 $A_m\overline{A}_{m+1}(m=n,n+1,\cdots)$ 及 A 互不相交，故由定义 2 中的 3° 得

$$P(A_n) = \sum_{m=n}^{\infty} P(A_m \overline{A}_{m+1}) + P(A).$$

仍然根据互不相交性，$\sum\limits_{m=1}^{\infty} P(A_m\overline{A}_{m+1}) \leqslant P(\Omega) = 1$, 故 (上) 式右方第一项是收敛级数的尾项，从而

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m \overline{A}_{m+1}) = 0.$$

由此即得 (15) . □

系 3 设 $A_n \in \mathcal{F}$, $A_n \subset A_{n+1}$, $n=1,2,\cdots$, 令 $A=\bigcup\limits_{n=1}^{\infty} A_n$, 则

$$P(A)=\lim_{n\to\infty} P(A_n). \tag{16}$$

证 集列 $\{\overline{A}_n\}$ 满足定理 2 中的条件，故如令 $\widetilde{A}=\bigcap\limits_{n=1}^{\infty} \overline{A}_n$, 则 $P(\widetilde{A})=\lim\limits_{n\to\infty} P(\overline{A}_n)$, 然而由 (7) 知

$$\widetilde{A}=\bigcap_{n=1}^{\infty} \overline{A}_n=\overline{\bigcup_{n=1}^{\infty} A_n}=\overline{A}.$$

故由 (13) 得

$$\begin{aligned} P(A) &= 1-P(\overline{A})=1-P(\widetilde{A})=1-\lim_{n\to\infty} P(\overline{A}_n) \\ &= \lim_{n\to\infty}[1-P(\overline{A}_n)]=\lim_{n\to\infty} P(A_n). \ \square \end{aligned}$$

最后我们指出，等式 (1) 与等式 (8) 是有很大区别的. 前者表示对任意可列多个互不相容的事件，概率是可加的，故 (1) 式所表达的性质称为完全可加性；(8) 式中表示对任意有穷多个互不相容的事件概率可加，故称 (8) 式所表达的性质为可加性. 如上所证，由完全可加性可以推出可加性；反面的结论一般是不成立的. 然而我们有下列定理：

***定理 3** 设 P 为 $\sigma-$ 代数 $\mathcal{F}$ 上的任意非负集函数，$P(\Omega)=1$. 则 P 具有完全可加性的充分与必要条件是：

(a) 它是可加的；

(b) 它是连续的：即对任一列 $A_n \in \mathcal{F}$, $A_n \supset A_{n+1}$, $n=1,2,\cdots$, 令 $A=\bigcap\limits_{n=1}^{\infty} A_n$, 则 (15) 式成立.

证 必要性已在定理 1 及定理 2 中证明；下面由 (a)(b) 证充分性. 设 $B_n \in \mathcal{F}$, $B_iB_j=\emptyset$, 又 $B=\bigcup\limits_{m=1}^{\infty} B_m$. 要证 $P(B)=\sum\limits_{m=1}^{\infty} P(B_m)$. 设 $A_n=\bigcup\limits_{m=n}^{\infty} B_m$, 显见 $A_n \supset A_{n+1}$, 令 $A=\bigcap\limits_{n=1}^{\infty} A_n$, 如果存在 $\omega \in A$, 那么 $\omega \in A_n=\bigcup\limits_{m=n}^{\infty} B_m$, 一切 n, 于是 ω 必属于无穷多个 B_m. 这与假设 B_m 互不相交矛盾，因而得证 $A=\emptyset$; 根据 (b) 知 $\lim\limits_{n\to\infty} P(A_n)=P(A)=0$.

由于 $B=\bigcup\limits_{m=1}^{\infty} B_m \cup A_{n+1}$ 及 (a), 得

$$P(B)=\sum_{m=1}^{n} P(B_m)+P(A_{n+1}).$$

令 $n\to\infty$ 即得 $P(B)=\sum\limits_{m=1}^{\infty} P(B_m)$. $\square$

§1.4 条 件 概 率

(一) 条件概率的定义. 在实际问题中，除了要知道事件 A 的概率 $P(A)$ 外，有时还需要知道在“事件 B 已出现”的条件下，事件 A 出现的条件概率 $P(A|B)$. 由于增加了新的条件：“事件 B 已出现”, 所以 $P(A)$ 一般是与 $P(A|B)$ 不同的.

例如，从标号为 1 、 2 、 3 、 4 的四个球中，等可能地任取一球，那么事件 A: “得标号为 4” 的概率 $P(A)=\frac{1}{4}$; 如果已知事件 B:“得标号为偶数” 已出现，那么这时只剩下两种可能，或得 2 号或得 4 号，所以 $P(A|B)=\frac{1}{2}$.

在一般情形下，应该怎样定义 $P(A|B)$? 由于频率与概率有很多类似的性质，先从频率的讨论开始.

设 A,B 为任一随机试验 E 中的两个事件，每次试验结果，不外是下列四种情形之一.

(1) A 出现， B 不出现； (2) B 出现， A 不出现；

(2) A,B 都出现； (4) A,B 都不出现.

现在把 E 重复做 n 次，分别以 n_1,n_2,n_3,n_4 记四种情况出现的次数，显然 $\sum\limits_{i=1}^{4} n_i=n$. 而且

$$B\text{的频率为}F_n(B)=\frac{n_2+n_3}{n},$$
$$AB\text{的频率为}F_n(AB)=\frac{n_3}{n}.$$

在 B 已出现的条件下， A 的频率为 $F_n(A|B)=\frac{n_3}{n_2+n_3}$. 根据这些式子，得

$$F_n(AB)=F_n(B)\cdot F_n(A|B). \tag{1}$$

因此，如 $F_n(B)>0$, 就有

$$F_n(A|B)=\frac{F_n(AB)}{F_n(B)}. \tag{2}$$

(2) 式启发我们应如何定义 $P(A|B)$.

定义 1 设 $(\Omega,\mathcal{F},P)$ 为概率空间， $A\in\mathcal{F}$, $B\in\mathcal{F}$, 设 $P(B)>0$. 在事件 B 已出现的条件下，事件 A 出现的条件概率 $P(A|B)$ 定义为

$$P(A|B)=\frac{P(AB)}{P(B)}. \tag{3}$$

对于古典型随机试验，设 B 含 m 个不同的基本事件， $m>0$, AB 含 k 个，以 n 表 Ω 中总共不同的基本事件的个数，则

$$P(A|B)=\frac{k/n}{m/n}=\frac{k}{m}.$$

类似地，对几何型随机试验，如 $L(B)>0$, 我们有

$$P(A|B)=\frac{L(AB)/L(\Omega)}{L(B)/L(\Omega)}=\frac{L(AB)}{L(B)}.$$

定理 1 如果 $P(B)>0$, 那么 $P(A|B)$ 作为 A 的集函数是 $\mathcal{F}$ 上的概率；即

(i) 对每 $A\in\mathcal{F}$, 有 $1\geqslant P(A|B)\geqslant 0$;

(ii) $P(\Omega|B)=1$;

(iii) 如 $A_m\in\mathcal{F}$, $m=1,2,\cdots$, $A_iA_j=\emptyset$, $i\neq j$, 则

$$P\left(\bigcup_{m=1}^{\infty} A_m|B\right)=\sum_{m=1}^{\infty} P(A_m|B). \tag{4}$$

证 (i) 因 $P(B) \geqslant P(AB)$, $P(B) > 0$, 故由 (3) 知 $1 \geqslant P(A|B) \geqslant 0$.

(ii) $$P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

(iii) $$P\left(\bigcup_{m=1}^{\infty} A_m|B\right) = \frac{P\left(\bigcup\limits_{m=1}^{\infty} A_m B\right)}{P(B)} = \sum_{m=1}^{\infty} \frac{P(A_m B)}{P(B)} = \sum_{m=1}^{\infty} P(A_m|B). \quad \square$$

既然条件概率也是一种概率，故对概率所证明的结果都适用于条件概率.

(二) 有关条件概率的三定理. 现在对条件概率来证明三条重要的定理，这就是：概率的乘法定理，全概率公式与巴叶斯 (Bayes) 公式. 这些定理在概率的计算中起着重要的作用.

定理 2 设 $A_1, A_2, \cdots, A_n$ 为 n 个事件， $n \geqslant 2$, 满足 $P(A_1A_2\cdots A_{n-1}) > 0$; 则

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{5}$$

(5) 式称为乘法公式. 它的直观意义是： $A_1, A_2, \cdots, A_n$ 同时出现的概率，等于先出现 A_1, 在 A_1 出现的条件下出现 A_2, 在 A_1, A_2 出现的条件下出现 $A_3, \cdots$ 各自的概率的乘积.

证 由于 $P(A_1) \geqslant P(A_1A_2) \geqslant \cdots \geqslant P(A_1A_2\cdots A_{n-1}) > 0$, 故 (5) 式右方出现的条件概率都有意义；由条件概率的定义， (5) 式右方等于

$$P(A_1)\frac{P(A_1A_2)}{P(A_1)}\frac{P(A_1A_2A_3)}{P(A_1A_2)}\cdots\frac{P(A_1A_2\cdots A_n)}{P(A_1A_2\cdots A_{n-1})} = P(A_1A_2\cdots A_n). \quad \square$$

例 1 设箱内有 $a(a \geqslant 2)$ 个白球 b 个黑球，在其中接连取三次，每次取一球，取后不还原，问三个都是白球的概率是多少？

解 以 A_i 表 “第 i 次取得白球” 这一事件， $i = 1, 2, 3$, 要求的是 $P(A_1A_2A_3)$. 因为

$$P(A_1A_2) = \frac{\binom{a}{2}}{\binom{a+b}{2}} > 0,$$

故可用 (5). 显然 $P(A_1) = \frac{a}{a+b}$. 如已知第一次取得白球，箱内只剩下 $a-1$ 个白球 b 个黑球，可见 $P(A_2|A_1) = \frac{a-1}{(a-1)+b}$, 类似得 $P(A_3|A_1A_2) = \frac{a-2}{(a-2)+b}$. 于是由 (5) 得

$$P(A_1A_2A_3) = \frac{a}{a+b} \cdot \frac{a-1}{a+b-1} \cdot \frac{a-2}{a+b-2}.$$

* 注. 例 1 中的随机试验 $\widetilde{E}$ 是复合的： $\widetilde{E} = E_1 \times E_2 \times E_3$. Ω_1 共含 $a+b$ 个 ω_1, Ω_2 共含 $a+b-1$ 个 ω_2, Ω_3 共含 $a+b-2$ 个 ω_3, $A_1 =$(白球，球，球), $A_2 =$(球，白球，球), $A_3 =$(球，球，白球), 这里 “球” 不论是白或黑均可. 事件 A_1 对第一次试验的结果加了条件， $P(A_1) = \frac{a}{a+b}$. 如已知 A_1 出现，那么 Ω_2 由 $a-1$ 个白球 b 个黑球构成，故 $P(A_2|A_1) = \frac{a-1}{(a-1)+b}$. 如已知前二次都得白球，则 Ω_3 由 $a-2$ 个白球 b 个黑球构成，故 $P(A_3|A_1A_2) = \frac{a-2}{a+b-2}$. 注意随机试验 E_2 依赖于 E_1 试验后的结果， E_3 依赖于 E_1 及 E_2 的结果，故 E_1, E_2, E_3 是相依的随机试验.

定理 3 设 $H_1, H_2, \cdots$ 为有穷或可列多个互不相容的事件， $P(\bigcup\limits_n H_n) = 1$, $P(H_n) > 0$, $(n = 1, 2, \cdots)$, 则对任一事件，有

$$P(A) = \sum_n P(H_n)P(A|H_n). \tag{6}$$

(6) 式称为全概率公式.

证 由 $P(\bigcup_n H_n)=1$ 得 $P(\overline{(\bigcup_n H_n)})=0$. 因为 H_n 互不相容，故 AH_n 也互不相容，$n=1,2,\cdots$, 于是

$$\begin{aligned}P(A)=P(A\Omega)&=P\Big(A\Big(\bigcup_n H_n\Big)\Big)+P\Big(A\overline{\Big(\bigcup_n H_n\Big)}\Big)\\&=P\Big(\bigcup_n AH_n\Big)=\sum_n P(AH_n).\end{aligned}$$

由 (5),$P(AH_n)=P(H_n)P(A|H_n)$; 代入上式即得 (6) . □

例 2 设甲箱中有 a 个白球 b 个黑球， $a>0,b>0$, 乙箱中有 c 个白球 d 个黑球，自甲箱中任意取一球放入乙箱，然后再从乙箱中任意取一球，试求事件 A:“从乙箱中取得的球为白球” 的概率.

解 以 $H_1(H_2)$ 表事件 “自甲箱中取出的球为白 (黑) 球”, 显然 $H_1H_2=\emptyset, H_1\cup H_2=\Omega$, 故 $P(H_1\cup H_2)=1$, 又 $P(H_1)=\frac{a}{a+b}>0$, $P(H_2)=\frac{b}{a+b}>0$, 由全概率公式

$$P(A)=\sum_{n=1}^{2}P(H_n)P(A|H_n). \tag{7}$$

但如 H_1 出现，那么乙箱中有 $c+1$ 个白球， d 个黑球，故 $P(A|H_1)=\frac{c+1}{c+d+1}$; 类似得 $P(A|H_2)=\frac{c}{c+d+1}$. 代入 (7) 中便得

$$\begin{aligned}P(A)&=\frac{a}{a+b}\frac{c+1}{c+d+1}+\frac{b}{a+b}\frac{c}{c+d+1}\\&=\frac{ac+bc+a}{(a+b)(c+d+1)}.\end{aligned}$$

定理 4 设 $H_1,H_2,\cdots$ 为有穷或可列多个互不相容的事件， $P(\bigcup_n H_n)=1$, $P(H_n)>0$, $(n=1,2,\cdots)$, 则对任一事件 A, $P(A)>0$, 有

$$P(H_m|A)=\frac{P(A|H_m)P(H_m)}{\sum_n P(A|H_n)P(H_n)}. \tag{8}$$

(8) 式称为巴叶斯公式.

证 由条件概率的定义及全概率公式得

$$P(H_m|A)=\frac{P(AH_m)}{P(A)}=\frac{P(A|H_m)P(H_m)}{\sum_n P(A|H_n)P(H_n)}. \quad \square$$

例 3 设甲乙丙三个箱子中

甲箱内有 a_1 个白球 b_1 个黑球,

乙箱内有 a_2 个白球 b_2 个黑球,

丙箱内有 a_3 个白球 b_3 个黑球，$(a_1+a_2+a_3>0)$

今任意取出一箱，再自此箱中任意取出一球，结果发现此球为白球．试在事件 A“此球为白球”的条件下，求事件 H_1“此球属于甲箱”的条件概率 $P(H_1|A)$.

解 这里 $P(H_1)=P(H_2)=P(H_3)=\frac{1}{3}>0$, H_1,H_2,H_3 分别表“此球属于甲箱”“乙箱”“丙箱”, 这三事件互不相容，$\bigcup_{n=1}^{3}H_n=\Omega$, 故 $P\left(\bigcup_{n=1}^{3}H_n\right)=1$; 又由全概率公式

$$\begin{aligned}P(A)&=\sum_{n=1}^{3}P(H_n)P(A|H_n)\\&=\frac{1}{3}\frac{a_1}{a_1+b_1}+\frac{1}{3}\frac{a_2}{a_2+b_2}+\frac{1}{3}\frac{a_3}{a_3+b_3}>0.\end{aligned}$$

故可用巴叶斯公式得

$$\begin{aligned}P(H_1|A)&=\frac{\frac{1}{3}\cdot\frac{a_1}{a_1+b_1}}{\frac{1}{3}\frac{a_1}{a_1+b_1}+\frac{1}{3}\frac{a_2}{a_2+b_2}+\frac{1}{3}\frac{a_3}{a_3+b_3}}\\&=1\Big/\left[1+\frac{a_2(a_1+b_1)}{a_1(a_2+b_2)}+\frac{a_3(a_1+b_1)}{a_1(a_3+b_3)}\right].\ \square\end{aligned}$$

巴叶斯公式通常用在下列实际问题中：设只可能出现 $H_1,H_2,\cdots$ 共有有穷或可列多种不同的情况，而事件 A 只能伴随这些情况之一发生．今在 A 已出现的条件下，试求发生了情况 H_m 的条件概率．

例 4 有朋友自远方来访，他乘火车来的概率是 3/10, 乘船，或乘汽车，或乘飞机来的概率分别为 1/5,1/10,2/5, 如果他乘火车来，迟到的概率是 1/4; 如果乘船‘或乘汽车，那么迟到的概率分别为 1/3,1/12; 如果乘飞机便不会迟到 (因而．这时迟到的概率为 0). 结果他是迟到了，试问在此条件下，他是乘火车来的概率等于多少．

解 以事件 A 表“迟到”,H_1,H_2,H_3,H_4 分别表“乘火车”“乘船”“乘汽车”“乘飞机”, 于是

$$\begin{aligned}P(H_1|A)&=\frac{P(H_1)P(A|H_1)}{P(H_1)P(A|H_1)+P(H_2)P(A|H_2)+P(H_3)P(A|H_3)+P(H_4)P(A|H_4)}\\&=\frac{\frac{3}{10}\cdot\frac{1}{4}}{\frac{3}{10}\cdot\frac{1}{4}+\frac{1}{5}\cdot\frac{1}{3}+\frac{1}{10}\cdot\frac{1}{12}+\frac{2}{5}\cdot 0}=\frac{1}{2}.\end{aligned}$$

注意 $P(H_1|A)=\frac{1}{2}$ 与 $P(H_1)=\frac{3}{10}$ 是不同的．类似地，如果以 A 的对立事件 $\overline{A}$(不迟到) 代替上式中的 A, 就得

$$P(H_1|\overline{A})=\frac{\frac{3}{10}\cdot\frac{3}{4}}{\frac{3}{10}\cdot\frac{3}{4}+\frac{1}{5}\cdot\frac{2}{3}+\frac{1}{10}\cdot\frac{11}{12}+\frac{2}{5}\cdot 1}=\frac{9}{34}.\ \square$$

*(三) **较复杂的例**．下列各例可以说明上述定理的联合应用．

例 5 设甲乙二人自 a 个白球 b 个黑球中任取一球，从甲开始然后轮流取，每次取后不还原，试求甲 (或乙) 先取得白球的概率 p_1(或 p_2).

解 为了使甲先取出一白球，必须也只须或者甲第一次就取得白球 (简记为“白”)，或者甲第一次取得黑球，乙第二次取得黑球，甲第三次取得白球 (简记为“黑、黑、白”)……因而，事件“甲先得白球”可表为互不相容的事件“白”、“黑、黑、白”、“黑、黑、黑、黑、

白”…… 的和，然而事件“白”的概率为 $\frac{a}{a+b}$，事件“黑、黑、白”的概率可由 (5) 式算得出 $\frac{b}{a+b}\cdot\frac{b-1}{a+b-1}\cdot\frac{a}{a+b-2}$，事件“黑、黑、黑、黑、白”的概率仍由 (5)，$\frac{b}{a+b}\cdot\frac{b-1}{a+b-1}\cdot\frac{b-2}{a+b-2}\cdot\frac{b-3}{a+b-3}\cdot\frac{a}{a+b-4}$，…，所以

$$p_1=\frac{a}{a+b}\left[1+\frac{b(b-1)}{(a+b-1)(a+b-2)}+\frac{b(b-1)(b-2)(b-3)}{(a+b-1)(a+b-2)(a+b-3)(a+b-4)}+\cdots\right] \tag{9}$$

同样得

$$p_2=\frac{a}{a+b}\left[\frac{b}{a+b-1}+\frac{b(b-1)(b-2)}{(a+b-1)(a+b-2)(a+b-3)}+\cdots\right].\ \square \tag{10}$$

注意，由于 b 是有穷数，故上二式右方中自某一项起全为 0; 又因为甲、乙二人中，总有一人先取出白球，故 $p_1+p_2=1$. 以 p_1,p_2 的值代入并简化后，立得等式

$$1+\frac{b}{a+b-1}+\frac{b(b-1)}{(a+b-1)(a+b-2)}+\cdots=\frac{a+b}{a}. \tag{11}$$

于是我们附带地用概率的方法证明了恒等式 (11). 用概率的方法来证明一些关系式或解决其它一些数学分析中的问题，是概率论的重要研究方向之一.

例 6 自 a 个白球 b 个黑球中同时任取 n 个球，$a+b\geqslant n$，试求至少取出一白球的概率 p.

解 先求对立事件的概率，事件 B:“取出的全是黑球”的概率是

$$q=\frac{\binom{b}{n}}{\binom{a+b}{n}}\ \left[\text{回忆}\ \binom{b}{n}=0,\ \text{如}\ b<n\right],$$

故

$$p=1-q=1-\frac{b(b-1)\cdots(b-n+1)}{(a+b)(a+b-1)\cdots(a+b-n+1)}. \tag{12}$$

还可以用另一个方法求出 p: 同时取出 n 个球可看成不还原地连取 n 次，每次取一球. 为了使 n 次中至少取出一白球，必须也只须或者第一次就得白球 (概率为 $\frac{a}{a+b}$)，或者第一次得黑球第二次得白球 (概率为 $\frac{b}{a+b}\frac{a}{a+b-1}$)，……，这些事件互不相容，故

$$p=\frac{a}{a+b}+\frac{b}{a+b}\frac{a}{a+b-1}+\cdots+\frac{b}{a+b}\frac{b-1}{a+b-1}\cdot\cdots\cdot\frac{b-n+2}{a+b-n+2}\frac{a}{a+b-n+1}. \tag{13}$$

比较 (12)(13)，可见它们右方的值相等，于是又得到恒等式：当 $a>0$ 时

$$\begin{aligned}1&+\frac{b}{a+b-1}+\cdots+\frac{b(b-1)\cdots(b-n+2)}{(a+b-1)(a+b-2)\cdots(a+b-n+1)}\\&=\frac{a+b}{a}\left[1-\frac{b(b-1)\cdots(b-n+1)}{(a+b)(a+b-1)\cdots(a+b-n+1)}\right].\ \square\end{aligned} \tag{14}$$

下面的例 7 是个著名的问题，于 1708 年为蒙特摩特 (Montmort) 所解决，后由拉普拉斯 (Laplace) 等人所推广.

例 7 (配对问题) 有 n 张信纸，分别标号为 $1,2,\cdots,n$，另有 n 个信封也同样标号，今将每张信纸任意装入一信封，试求“没有一个配对”的概率 q_0 及“恰有 r 个配对”的概率 q_r，$(r\leqslant n)$，这里所谓“r 个配对”是指有 r 张信纸，分别装入同号码的信封.

解 以 A_i 表"第 i 号信纸装入第 i 号信封"这一事件，则

$$q_0 = 1 - P\left(\bigcup_{i=1}^{n} A_i\right).$$

为求 $P\left(\bigcup_{i=1}^{n} A_i\right)$, 利用一般加法公式 (§1.3 系 2). 第 i 号信纸可装入 n 个信封，恰好装入第 i 号信封的概率 $P(A_i) = \frac{1}{n}$, 故

$$s_1 = \sum_i P(A_i) = 1.$$

如 A_i 出现，第 j 号信纸共有 $n-1$ 个信封可以选择，故

$$P(A_j|A_i) = \frac{1}{n-1}, \ P(A_iA_j) = P(A_i)P(A_j|A_i) = \frac{1}{n} \cdot \frac{1}{n-1},$$

从而

$$s_2 = \sum_{i,j} P(A_iA_j) = \binom{n}{2} \Big/ n(n-1) = \frac{1}{2!}.$$

类似地一般有

$$s_r = \frac{1}{r!}, \quad (r = 1, 2, \cdots, n).$$

于是

$$q_0 = 1 - P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

注意 q_0 与 n 有关，如记 $q_0 = q_0(n)$, 则

$$\lim_{n\to\infty} q_0(n) = e^{-1} = 0.36\cdots.$$

利用 q_0 便不难求出 q_r. 如果指定某 r 张信纸装入对应的信封，这事件的概率为

$$\frac{1}{n(n-1)\cdots(n-r+1)};$$

其余 $n-r$ 张信纸中没有一个配对的概率为

$$q_0(n-r) = \sum_{k=0}^{n-r} \frac{(-1)^k}{k!};$$

由于 r 张配对的信纸共有 $\binom{n}{r}$ 种选法，故

$$q_r = \frac{\binom{n}{r}}{n(n-1)\cdots(n-r+1)} \cdot \sum_{k=0}^{n-r} \frac{(-1)^k}{k!} = \frac{1}{r!} \sum_{k=0}^{n-r} \frac{(-1)^k}{k!}.$$

注意当 $n \to \infty$ 时， $q_r = q_r(n) \to e^{-1}/r!$. □

配对问题具有典型的意义，例如"信纸"可设想为"旅客"或"机床","信封"可设想为他们的"箱子"或"零件"等等 (参看 101 页).

§1.5 独　立　性

(一) **二事件的独立性**. 设 A,B 为二事件，如果 $P(B)>0$, 可以定义 $P(A|B)$. 一般说来， $P(A)\neq P(A|B)$. 直观地，这表示 B 的出现对 A 出现的概率是有影响的，只是当 $P(A)=P(A|B)$ 时，才可以认为这种影响不存在，这时自然会设想 A,B 是彼此独立的. 由 §1.4(5), 如果 $P(A)=P(A|B)$, 就有

$$P(AB)=P(A)P(B). \tag{1}$$

这引出下列定义:

定义 1 如果二事件 A,B 满足 (1) 式，则称 A,B 是相互独立的.

注意，定义 1 即使在 $P(A)=0$ 或 $P(B)=0$ 的情况下仍然适用.

例 1 分别掷二枚硬币. 事件 A:“甲出现正面” 与事件 B:“乙出现正面” 是相互独立的，实际上，这时 Ω =((正，正),(正，反),(反，正),(反，反)) 共含四个基本事件，它们是等可能的，各有概率 $\frac{1}{4}$. 显然:

$$A=((\text{正},\ \text{正}),(\text{正},\ \text{反})),B=((\text{正},\ \text{正}),(\text{反},\ \text{正})),AB=((\text{正},\ \text{正}))$$

从而

$$P(AB)=\frac{1}{4}=\frac{1}{2}\cdot\frac{1}{2}=P(A)P(B).\ \square$$

不难看出: 如果 $P(B)>0$, 那么 A,B 的相互独立性等价于

$$P(A)=P(A|B).$$

实际上，上面已证明由上式可得 (1), 反之，如 (1) 成立，由 (1) 及 §1.4(5) 得

$$P(A)P(B)=P(AB)=P(B)P(A|B).$$

以 $P(B)(>0)$ 除此式两边，就得上式.

定理 1 如果 (A,B) 相互独立，那么三对事件 $(A,\overline{B}),(\overline{A},B),(\overline{A},\overline{B})$ 分别也是相互独立的.

证　由 $A=AB\cup A\overline{B}$ 得

$$P(A)=P(AB)+P(A\overline{B}).$$

如 $P(AB)=P(A)P(B)$, 则

$$P(A\overline{B})=P(A)-P(A)P(B)=P(A)[1-P(B)]=P(A)P(\overline{B}).$$

故 $(A,\overline{B})$ 相互独立. 由 A,B 的对称性可见此时 $(\overline{A},B)$ 也相互独立. 以所证得的结果用于 $(A,\overline{B})$, 可见 $(\overline{A},\overline{B})$ 相互独立.

*(二) **n 个事件的独立性**.

定义 2 设 $A_1, A_2, \cdots, A_n$ 是 n 个事件，我们说这些事件是相互独立的，如果对任意 $s(2 \leqslant s \leqslant n)$, 任意 $1 \leqslant i_1 < i_2 < \cdots < i_s \leqslant n$, 有

$$P(A_{i_1} A_{i_2} \cdots A_{i_s}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_s}). \tag{2}$$

注意，(2) 式共代表 $2^n - n - 1$ 个等式，实际上

$$\left.\begin{aligned} &P(A_{i_1} A_{i_2}) = P(A_{i_1}) P(A_{i_2}) \quad (\text{共}\binom{n}{2}\text{个}), \\ &P(A_{i_1} A_{i_2} A_{i_3}) = P(A_{i_1}) P(A_{i_2}) P(A_{i_3}) \quad (\text{共}\tbinom{n}{3}\text{个}), \\ &\qquad \cdots\cdots, \\ &P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n) \quad (\text{共}\binom{n}{n}\text{个}). \end{aligned}\right\} \tag{3}$$

故总共的个数为

$$\sum_{i=2}^{n} \binom{n}{i} = (1+1)^n - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1.$$

由定义 2 立刻看出，如果 $A_1, A_2, \cdots, A_n$ 相互独立，那么其中任意 m 个事件也相互独立，$m \leqslant n$.

* 现在来推广定理 1. 考虑 (3) 中最后一式

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n)$$

如果在 n 个事件中，任取 m 个，$0 \leqslant m \leqslant n$, 并把这 m 个事件 A_j 换成它的对立事件 $\overline{A}_j$, 然后用 $\overline{A}_j$ 代替上述最后一式中的 A_j, 于是就得到 2^n 个式子：

$$\left.\begin{aligned} &P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n)\ (\text{共}\tbinom{n}{0}\text{个}), \\ &P(A_1 \cdots \overline{A}_i \cdots A_n) = P(A_1) \cdots P(\overline{A}_i) \cdots P(A_n)(\text{共}\tbinom{n}{1}\text{个}), \\ &P(A_1 \cdots \overline{A}_i \cdots \overline{A}_j \cdots A_n) = P(A_1) \cdots P(\overline{A}_i) \cdots P(\overline{A}_j) \cdots P(A_n)(\text{共}\tbinom{n}{2}\text{个}), \\ &\qquad \cdots\cdots, \\ &P(\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n) = P(\overline{A}_1) P(\overline{A}_2) \cdots P(\overline{A}_n) \quad (\text{共}\tbinom{n}{n}\text{个}). \end{aligned}\right\} \tag{4}$$

***定理 2** [1] 为使事件 $A_1, A_2, \cdots, A_n$ 相互独立，必要与充分条件是 (4) 中 2^n 个等式成立.

证 充分性：由 (4) 得

$$P(A_1 \cdots, A_{n-1} \overline{A}_n) = P(A_1) \cdots P(A_{n-1}) P(\overline{A}_n).$$

将这式与 (4) 中第一式相加便得

$$P(A_1 \cdots, A_{n-1}) = P(A_1) \cdots P(A_{n-1}).$$

这便得到了 (3) 中所需要证明的一个等式. (3) 中其它等式可同样地证明.

[1] 定理 2 是 §2.7 定理 1 的特殊情形.

必要性：(3) 中最后一式与 (4) 中第一式重合，试证 (4) 中第二式. 由 (3) 得

$$P(A_1\cdots A_{i-1}A_{i+1}\cdots A_n\cdot A_i)=P(A_1\cdots A_{i-1}A_{i+1}\cdots A_n)P(A_i).$$

这表示二事件 $A_1\cdots A_{i-1}A_{i+1}\cdots A_n$, A_i 相互独立，由定理 1 知 $A_1\cdots A_{i-1}A_{i+1}\cdots A_n$ 与 $\overline{A}_i$ 也相互独立，再由 (3) 便得

$$\begin{aligned}P(A_1\cdots A_{i-1}A_{i+1}\cdots A_n\cdot\overline{A}_i)&=P(A_1\cdots A_{i-1}A_{i+1}\cdots A_n)P(\overline{A}_i)\\&=P(A_1)\cdots P(\overline{A}_i)\cdots P(A_n).\end{aligned}$$

于是 (4) 中第二式得证，即得证

$$P(A_1\cdots\overline{A}_i\cdots A_n)=P(A_1)\cdots P(\overline{A}_i)\cdots P(A_n).\tag{5}$$

由 (5) 及 (3), 可见 $A_1\cdots\overline{A}_i\cdots A_{j-1}A_{j+1}\cdots A_n$ 与 A_j 相互独立，于是用同样的方法，可证 (4) 中第三式正确，这样类推下去，便可完全证明 (4) . □

***系 1** 设事件 $(A_1,A_2,\cdots,A_n)$ 相互独立，则事件 $(\overline{A}_{i_1},\overline{A}_{i_2},\cdots,\overline{A}_{i_m},A_{i_{m+1}},\cdots,A_{i_n})$ 也相互独立，其中 $\{i_1,i_2,\cdots,i_n\}$ 是 $\{1,2,\cdots,n\}$ 的任一排列，$1\leqslant m\leqslant n$.

证 由假设及定理 2, 对 $A_1,A_2,\cdots,A_n$,(4) 式成立，对 $\overline{A}_{i_1},\overline{A}_{i_2},\cdots,\overline{A}_{i_m},A_{i_{m+1}},\cdots,A_{i_n}$ 的 (4) 式记为 (4)', 注意 $\overline{\overline{A}}=A$ 后可见 (4) 与 (4)' 重合，故 (4)' 也成立，再由定理 2 即得所欲证. □

在独立性概念中，必须注意下列事实：由事件 $A_1,A_2,\cdots,A_n$ 的两两独立性不能推出它们的相互独立性. 所谓两两独立性只是指 (3) 中第一式 (只含 $\binom{n}{2}$ 个等式) 成立，这由下列例 2 可见.

例 2 设 $\Omega=(\omega_1,\omega_2,\omega_3,\omega_4)$ 是等可能的，即 $P(\omega_i)=\frac{1}{4}$ $(i=1,2,3,4)$. 令

$$A=(\omega_1,\omega_2),\ B=(\omega_1,\omega_3),\ C=(\omega_1,\omega_4).$$

则 $P(A)=P(B)=P(C)=\frac{1}{2}$. 由于 $AB=AC=BC=(\omega_1)$, 有

$$\begin{aligned}P(AB)=P(AC)=P(BC)&=\frac{1}{4}=\frac{1}{2}\cdot\frac{1}{2}\\&=P(A)P(B)=P(A)P(C)=P(B)P(C).\end{aligned}$$

因而三事件 A,B,C 是两两独立的. 然而它们却不相互独立. 因为

$$P(ABC)=P(\{\omega_1\})=\frac{1}{4}\neq\frac{1}{8}=P(A)P(B)P(C)$$

从而 (3) 中最后一式不成立. □

例 3 设一次射击中的的概率为 p, $0<p<1$, 不断地接连射下去，假定各次射击都独立地进行，试求事件 A_k:"第 k 次射击中的",B_k: "前 k 次射击里，至少中的一次",C: "迟早至少中的一次" 的概率.

解 以 1 表 "中的",0 表 "不中的". 由于射击不断地进行，故得到可列次重复随机试验 $\widetilde{E}=E^\infty$,(一次射击为 E). $\widetilde{E}$ 的一个基本事件是 $(x_1,x_2,\cdots)$, 其中 $x_i=1$ 或 0, 因而 $\widetilde{\Omega}$ 是非可列集，实际上 $\widetilde{\Omega}$ 与 $[0,1]$ 是一一对应的.

事件 A_k 由一切如下序列 $(x_1,\cdots x_k,\cdots)$ 组成：其中 $x_k=1$, 其余的 $x_i(i\neq k)$ 可任意为 0 或为 1. 故 A_k 只依赖于第 k 次射击的结果，$P(A_k)=p$.

B_k 只依赖于前 k 次射击. B_k 的对立事件 $\overline{B}_k$ 为："前 k 次射击都不中的", 显然

$$\overline{B}_k=\overline{A}_1\overline{A}_2\cdots\overline{A}_k.$$

由于各次射击是独立进行的，可以假定上式右方各事件独立，故如令 $q=1-p, 0<q<1$, 则

$$P(\overline{B}_k)=P(\overline{A}_1)P(\overline{A}_2)\cdots P(\overline{A}_k)=(1-p)^k=q^k. \tag{6}$$

于是 $P(B_k)=1-q^k$.

类似地有

$$\overline{C}=\bigcap_{k=1}^{\infty}\overline{B}_k.$$

因为 $\overline{B}_k\supset\overline{B}_{k+1}$, 故由 §1.3 定理 2

$$P(\overline{C})=\lim_{k\to\infty}P(\overline{B}_k)=\lim_{k\to\infty}q^k=0. \tag{7}$$

所以 $P(C)=1-P(\overline{C})=1$. □

独立事件常伴随独立随机试验列而出现：设 $\{E_i\}(i=1,2,\cdots)$ 是一列随机试验，E_i 的基本事件空间为 Ω_i, 设 A_k 是 E_k 中的任一事件，$A_k\subset\Omega_k$, 如果 A_k 出现的概率不依赖于其它各次试验 $E_i(i\neq k)$ 的试验结果，就称 $\{E_i\}$ 是独立随机试验列.

例 3 中的重复随机试验列 (不断地射击) 是独立的，而 §1.4 例 1 中的 $\{E_1,E_2,E_3\}$(不**还原** 抽球) 不是独立的，便还原抽球 (即每次把抽得的球还原后再独立地抽下次) 则一般地可以认为是独立随机试验列.

例 4 自 $1,2,\cdots,10$ 共十个数字中任取一个，取后还原，连取 k 次，独立进行，于是得到 k 个随机数字，试求事件 A_m:"此 k 个数字中最大者是 m" 的概率 $p_m(m\leqslant 10)$.

解 以 B_m 表事件："此 k 个数字中最大者不大于 m". 为了使 B_m 出现，必须也只须第一次取得的数字不大于 m(概率为 $\frac{m}{10}$) 第二次取得的也不大于 m(概率仍为 $\frac{m}{10}$),$\cdots$, 这样直到第 k 次. 由于随机试验列是独立的，故

$$P(B_m)=\left(\frac{m}{10}\right)^k. \tag{8}$$

容易看到，$B_m\supset B_{m-1}$, 而且

$$A_m=B_m-B_{m-1}.$$

故

$$P(A_m)=P(B_m)-P(B_{m-1})=\frac{m^k-(m-1)^k}{10^k}. \ \square \tag{9}$$

例 5 (小概率事件) 设随机试验 E 中某一事件 A 出现的概率为 $\varepsilon>0$. 试证不断独立地重复做试验 E 时，A 迟早会出现的概率为 1, 不论 ε 如何小.

证 以 A_k 表事件："A 于第 k 次试验中出现", $P(A_k)=\varepsilon$. 前 n 次试验中，A 都不出现的概率为

$$P(\overline{A}_1\overline{A}_2\cdots\overline{A}_n)=P(\overline{A}_1)P(\overline{A}_2)\cdots P(\overline{A}_n)=(1-\varepsilon)^n,$$

这里用到独立性. 于是前 n 次试验中， A 至少出现一次的概率为

$$1-P(\overline{A}_1\overline{A}_2\cdots\overline{A}_n)=1-(1-\varepsilon)^n\to 1,(n\to\infty).\ \square$$

这例说明决不能轻视小概率事件，尽管在一次试验中它出现的概率很小，但只要试验次数很大，而且是独立地进行的，那么它总会出现的概率就可以很接近于 1.

(三) **差分方程法**. 到现在为止，我们已讲过概率的定义及概率的一些重要性质，并列举了许多例题，后者显示了计算概率方法的多样性. 下面再给出一些例子，通过它们介绍另一种计算概率的方法 - 差分方程法 [1].

例 6 设质点 M 在整数点集 $(0,1,2,\cdots,a)$ 上作随机徘徊，就是说，每经一单位时间按下列规则改变一次位置：如果它现在在点 $z(0<z<a)$ 上. 下一步以概率 $p(0<p<1)$ 转移到 $z+1$, 以概率 q 到 $z-1$, $p+q=1$; 如果它现在在 0(或 a), 它以后就永远停留在 0(或 a) 试求自 z 出发，终于要到达 0(或 a) 的概率 q_z(或 p_z).

解 质点 M 自 $z(0<z<a)$ 出发，由于 $p+q=1$, 它下步来到 $z+1$ 或 $z-1$ 的概率为 1, 根据全概率公式 (见 §1.4 定理 3) 得

$$q_z=pq_{z+1}+qq_{z-1},\quad (0<z<a) \tag{10}$$

$$q_0=1,\quad q_a=0. \tag{11}$$

由此可见， $(q_0,q_1,\cdots,q_a)$ 是方程组 (10)(11) 的解. 下面证明 (10)(11) 的解唯一，并且具体求出此解. 由唯一性此解自然就是所要求的概率 $(q_0,q_1,\cdots,q_a)$.

首先注意 (10) 式有二特解为

$$q_z=1;\quad q_z=\left(\frac{q}{p}\right)^z \tag{12}$$

由 (10) 的齐次性，可见对任二常数 A,B,

$$q_z=A+B\left(\frac{q}{p}\right)^z \tag{13}$$

也是 (10) 的一解. 为使边值条件 (11) 满足，必须也只需

$$A+B=1,\quad A+B\left(\frac{q}{p}\right)^a=0,$$

由此二方程可以求出 A,B, 代入 (13) 即得

$$q_z=\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^2}{\left(\frac{q}{p}\right)^a-1}\ (\text{如}p\neq q) \tag{14}$$

是 (10)(11) 的解. 为证 (14) 是 (10)(11) 的唯一解，只要证 (13) 是 (10) 的解的一般形式. 设 $\{q_z\}$ 是 (10) 的任一解，在 (13) 中可选 A,B 使 $q_0=q_0',q_1=q_1'$; 然而由 (10) 知 q_{z+1} 被 q_{z-1},q_z 所唯一决定，从而由 $q_0=q_0'$ 及 $q_1=q_1'$ 知 $q_i=q_i'$, $i=0,1,2,\cdots$.

[1] 在另一些问题中还可用微分方程来求概率，参看 §4.3.

如 $p=q$, 上述讨论是行不通的，因为这时 (12) 中两特解重合. 然而此时可求得 (10) 的另一特解为 $q_z=z$, 因此 (10) 的通解仍可仿上证明为 $q_z=A+Bz$. 要边值条件 (11) 满足，必须令 $A=1, A+Ba=0$, 于是求得当 $p=q$ 时有

$$q_z=1-\frac{z}{a}. \tag{15}$$

类似的，为求 p_z, 注意它是下二方程

$$p_z=pp_{z+1}+qp_{z-1}\quad (0<z<a) \tag{16}$$

$$p_0=0,\quad p_a=1 \tag{17}$$

的解，用同样的解法，可以求得

$$p_z=1-q_z\quad (0\leqslant z\leqslant a).\ \square \tag{18}$$

在例 6 中，由于二边点 0,a, 都是吸引的，即 M 到达 0(或 a) 后，就永远停留在 0(或 a), 因此 q_z 可解释为质点 M 自 z 出发，在到达 a 以前先到达 0 的概率，而 p_z 是自 z 出发，在到 0 以前先到达 a 的概率. (18) 式还表示，它迟早总要被边点吸引，不可能永远在内部徘徊，因为 $p_z+q_z=1$.

在例 6 中，0,a 都是吸引的. 现在把 a 改为反射的，就是说：例 6 中所述的运动法则，除了当质点 M 自 a 出发时下一步以概率 1“停留在 a” 改为 “转移到 $a-1$” 外，其余都不改变. 试求自 z 出发，终于要达到 0 的概率 r_z. 这时，r_z 仍然满足差分方程 (10), 即

$$r_z=pr_{z+1}+qr_{z-1},\quad (0<z<a) \tag{19}$$

但边值条件 (11) 应换为

$$r_0=1,\quad r_a=r_{a-1}. \tag{20}$$

用类似的方法解此差分方程组 (19)(20), 即得

$$r_z=1\quad (0\leqslant z\leqslant a). \tag{21}$$

我们看到了，质点在边界点 0,a 上转移法则的改变对应于边值条件的改变.

有时直观的概率意义可以帮助我们检查所得结果是否正确. 例如，在例 6 中，容易想象，z 越近于 a, 质点被 0 吸引的可能性就越小，故 q_z 应是 z 的不增函数；(14)(15) 正有这个性质. 又如，若 $p=q$, $a=2n$, $z=n$ 是 $(0,2n)$ 的中点，易想到 $q_z=\frac{1}{2}$; 而 (15) 此时果然如此.

例 7 考虑随机试验 E 中一事件 A. $p=P(A)>0$. 将 E 独立地重复做下去，如果 A 接连出现至少 r 次，就说出现 “$r-$ 成功”(例如，$E-$ 射击，$A-$ 中的，$r-$ 成功就是接连中的至少 r 次), 试求在 n 次试验中出现 $r-$ 成功的概率 $y_n(n\geqslant r)$.

解 在 $n+1$ 次试验中，事件：“出现 $r-$ 成功” 可以分解为两个互不相容的事件；第一，在前 n 次试验中已出现 $r-$ 成功，概率为 y_n; 第二，只在 $n+1$ 次试验后才出现 $r-$ 成功，为计算第二事件的概率，注意此事件的出现等价于下列三事件 a,b,c 同时出现：

a: “在前 $n-r$ 次试验中 $r-$ 成功不出现”,$P(a)=1-y_{n-r}$;

b: "在第 $n-r+1$ 次试验中 A 不出现", $P(b)=q=1-p$;

c: "A 在其余 r 次试验中都出现", $P(c)=p^r$.

由于 a,b,c 是独立事件，故它们同时出现的概率 - 也就是上述第二事件的概率 - 等于 $(1-y_{n-r})qp^r$.

于是，得

$$y_{n+1}=y_n+(1-y_{n-r})qp^r \tag{22}$$

这是一个 $r+1$ 级的线性差分方程，连同显然的开始条件：

$$y_0=y_1=\cdots=y_{r-1}=0,\quad y_r=p^r,$$

可以完全决定 $y_n, n=r+1,r+2,\cdots$. 例如当 $r\geqslant 2$ 时

$$y_{r+1}=p^r+qp^r=p^r(1+q),$$
$$y_{r+2}=p^r+2qp^r=p^r(1+2q),\cdots.\ \square$$

例 8 (传球问题) r 个人相互传球，每传一次时，传球者等可能地传给其余 $r-1$ 人中之一，试求第 n 次传球时，此球由最初发球者传出的概率 p_n(发球那一次算作第 0 次).

解 以 q_n 表第 n 次传球时由某甲 (非发球者) 传出的概率，由等可能性，此概率对其余 $r-1$ 人都是相同的，于是

$$p_n+(r-1)q_n=1. \tag{23}$$

为使第 n 次球由发球者传出，充要条件是第 $n-1$ 次球由其余 $r-1$ 人中之一传出 (对应的概率为 $(r-1)q_{n-1}$), 并传给发球者 (对应的概率为 $\frac{1}{r-1}$). 故

$$p_n=(r-1)q_{n-1}\cdot\frac{1}{r-1}=q_{n-1},\quad (n\geqslant 1) \tag{24}$$

由 (23)(24) 得 $p_n=\frac{1-p_{n-1}}{r-1}$, 亦即

$$p_n-\frac{1}{r}=\frac{1}{1-r}\Big(p_{n-1}-\frac{1}{r}\Big)=\Big(\frac{1}{1-r}\Big)^n\Big(1-\frac{1}{r}\Big)\ \ (\text{注意} p_0=1).$$

最后得

$$p_n=\frac{1}{r}\Big\{1-\Big(\frac{1}{1-r}\Big)^{n-1}\Big\}.$$

由此可见：当 $n\to\infty$ 时 $p_n\to\frac{1}{r}$. 这个结果是可以直观地理解的. 但预料不到的是：当 n 为偶数时， p_n 略大于 "极限" 概率 $\frac{1}{r}$, n 为奇数时，则略小于 $\frac{1}{r}$. 不过，当 $r=2$ 时，$p_n=1$ 或 0, 视 n 为偶或奇而定，这当然是对的.

注意，由 (24), $\lim\limits_{n\to\infty}q_n=\lim\limits_{n\to\infty}p_n=\frac{1}{r}$. 这表明第 n 次发球的机会对所有 r 个人都是近似地等可能的，只要 n 充分大，而不管最初发球者是谁. 这种不依赖于开始状态的性质在马尔科夫链理论中叫做遍历性，其实本例就是一个简单的马尔科夫链 [1].

[1] 如利用 §4.1 中的理论，则本例中的转移概率 $p_{ij}=\frac{1}{r-1}, (i\neq j); p_{ii}=0$, 以此代入 §4.1(22) 式，即可求出那里的 $p_j=\frac{1}{r}, (j=1,\cdots,r)$, 即极限概率为 $\frac{1}{r}$.

例 8 的抽象化是：质点在 r 个位置上随机运动，每次位移从一位置到其余 $r-1$ 个位置是等可能的，试求第 n 次运动是由开始位置出发的概率 p_n.

*§1.6 若 干 补 充

(一) 本章小结.

1. 现实中许多随机现象可以用 (复合) 随机试验来概括. 我们着重地讨论了两种随机试验：古典型的与几何型的. 对这两种随机试验发现了概率的一些性质 (§1.1 定理 1,2), 而且发现一般的随机试验的频率也有类似性质 (§1.1 定理 3), 于是便以这些性质来定义 **一般** 的概率，从而建立了公理系统 (§1.3).

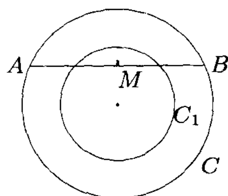
2. 我们已经看到计算古典型概率的三种方法：直接用 $P(A)=\frac{k}{n}$, 这时必须算好基本事件的个数 n 及 A 所含基本事件个数 k, 为此常常用到排列与组合. (2) 利用一些重要的公式或定理 (乘法公式，全概率公式，巴叶斯公式，连续性定理等), 见 §1.3 中的例及 §1.5 例 3 等等. (3) 差分方程 (§1.5).

3. 本章有许多重要的例，显示了问题的广泛性与方法的多样性. 读者可把它们总结一下，分成类 (例如抽球问题，分房问题，随机数问题，几何型概率问题等等), 或从方法上分类也可，并记住一些.

(二) 关于几何概率. 在 §1.1 中我们下了几何概率的定义 (见 §1.1(6)), 在运用这个定义来求具体问题中的概率时，必须注意所谓点的 "均匀分布"(它实际上是广泛意义下的等可能性) 是相对什么随机试验而言的，否则就可能得出不同的甚至错误的结果. 下面的例清楚地说明了这一点.

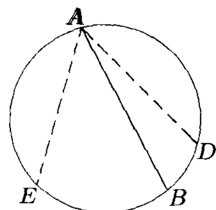
例 1 (Bertrand 奇论) 在一半径为 r 的圆 C 内 "任意" 作一弦，试求此弦长度 l 大于圆内接等边三角形的边长 $r\sqrt{3}$ 的概率 p.

解 1 作半径为 $\frac{r}{2}$ 的同心圆 C_1. 弦的中点 M 是 "任意" 的. 如 $M\in C_1$ 则 $l>r\sqrt{3}$ 故由 §1.1(6),p 等于二圆面积之比，即



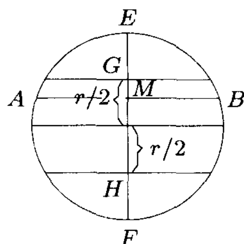
$$p=\frac{\pi\left(\frac{r}{2}\right)^2}{\pi r^2}=\frac{1}{4}.$$

解 2 由对称性不妨先固定弦的一端 A 于圆周上，于是另一端 B 是 "任意" 的. 考虑等边三角形 ΔADE, 如 B 落于角 A 所对应的弧 $\overset{\frown}{DE}$ 上，则 $l>r\sqrt{3}$. 故



$$p=\frac{\overset{\frown}{DE}\text{ 的弧长}}{\text{圆周全长}}=\frac{1}{3}.$$

解 3 不妨先固定弦的方向使它垂直于直径 EF. 如果 AB 的中点 M 落在右图的 GH 上，则 $l > r\sqrt{3}$. 因此



$$p = \frac{GH\text{之长}}{EF\text{之长}} = \frac{1}{2}.$$

于是我们得到了三个不同的答案，原因何在呢？这是因为三个解法中用了三个不同的随机试验的事件．第一个是“观察随机点 M 落于圆 C_1 中”(Ω 是二维区域 C), 第二个是“随机点 B 落于圆弧 $\overset{\frown}{ED}$ 上”(Ω 是全圆周), 第三个是“随机点 M 落于区间 GH 中”(Ω 是 EF). 因此三个解法中的 p 其实是三个不同事件的概率．这使我们看出，原来例 1 中的问题提得太不确定了，那里的“任意”二字至少可以作如上三种解释．相对于自己的解释，每种解法都是正确的． □

例 2 (最近星体的距离的分布) 在天体统计中，需要研究下列问题：设 A 为一固定星球，试求与 A 最邻近的星与 A 的距离不超过 x 的概率 $F(x)$.

解 如果不对在 A 周围星的密集程度作任何假定，问题是无法解决的．根据天文的观察，我们假设：

1) 以 A 为中心作一球 C, 体积为 V, 以 N 表位于 V 中的星数，则

$$\lim_{V\to\infty}\frac{N}{V} = \lambda, \ (\lambda\text{为某常数})$$

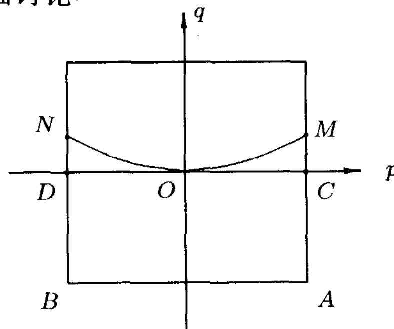
2) 星在 V 中相互独立地均匀分布．

以 v 表中心为 A 半径为 x 的球 S 的体积，由后一假设，C 中星都不在 S 中的概率为 $\left(1-\frac{v}{V}\right)^N$, 利用 1) 并注意 $v = \frac{4}{3}\pi x^3$, 得

$$F(x) = 1 - \lim_{V\to\infty}\left(1-\frac{v}{V}\right)^{\lambda V} = 1 - e^{-\frac{4}{3}\lambda\pi x^2}. \ \square \tag{1}$$

λ 随星球 A 而异．如 A 为太阳，可取 $\lambda = 0.0063$.

当 x 在 $[0,\infty)$ 变化时，$F(x)$ 是最近星的距离 r 的分布函数，后一概念在下章中要详细讨论．



例 3. 设二维点 (p,q) 在 $|p| \leqslant 1, |q| \leqslant 1$ 中按均匀分布出现，试求方程 $x^2+px+q=0$ 的二根

(i) 都是实数的概率 P_1;

(ii) 都是正数的概率 P_2.

解. (i) 基本事件空间 Ω 是以原点为中心，边长为 2 的正方形，即

$$\Omega = ((p,q): \ |p| \leqslant 1, |q| \leqslant 1).$$

为使二根为实，(p,q) 必须也只须满足 $p^2 \geqslant 4q$, 这条件决定区域 $AMNB$, MN 是抛物线段 $p^2 = 4q, |q| \leqslant 1$. 注意 MOC 的面积为 $\int_0^1 \frac{p^2}{4}\,dp = \frac{1}{12}$. 故

$$P_1 = \frac{L(AMNB)}{L(\Omega)} = \frac{2+2\cdot\frac{1}{12}}{4} = \frac{13}{24}.$$

(ii) 为使二根为正，(p,q) 除应满足 $p^2 \geqslant 4q$ 外，还要满足 $p<0, p<\sqrt{p^2-4q}<-p$; 亦即还要满足 $p<0, q>0$. 这三个条件决定区域 NOD, 它的面积为 $\frac{1}{12}$. 故 $P_2=\frac{1}{12}\div 4=\frac{1}{48}$. □

习　题

1. 设 A,B,C 表示三个随机事件，试将下列事件用 A,B,C 表示出来.

a) A 出现，B,C 不出现，　b) A,B 都出现，而 C 不出现，

c) 所有三个事件都出现，　d) 三个事件中至少一个出现，

e) 三个事件都不出现，　f) 不多于一个事件出现，

g) 不多于二个事件出现，　h) 三个事件中至少二个出现.

2. 下面两式表示 A,B 之间有什么包含关系？

a) $A\cap B=A$,　b) $A\cup B=A$.

3. 求证：

$(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$; $(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$.

4. 证明下列等式：

a) $\overline{\overline{A}}=A$.　b) $\overline{(A\cup B)}=\overline{A}\cap\overline{B}$.

c) $\overline{A\cap B}=\overline{A}\cup\overline{B}$.　d) $A-B=A\cap\overline{B}$.

e) $\overline{(\overline{A}\cap\overline{B})}=A\cup B$.　f) $\overline{(\overline{A}\cup\overline{B})}=A\cap B$.

5. 简化下列各式：

a) $(A\cup B)\cap(B\cup C)$.　b) $(A\cup B)\cap(A\cup\overline{B})$.

c) $(A\cup B)\cap(A\cup\overline{B})\cap(\overline{A}\cup B)$.

6. 设 A_n, $n=1,2,\cdots$, 是随机事件，试证：

a) $P(A_1\cup\cdots\cup A_n)=1-P(\overline{A}_1\cap\cdots\cap\overline{A}_n)$.

b) $P\left(\bigcup_n A_n\right)=P(A_1)+P(A_2\cap\overline{A}_1)+P(A_3\cap\overline{A}_1\cap\overline{A}_2)+\cdots$.

7. 一部四卷的文集，按任意次序放到书架上，问各卷自左向右或自右向左的卷号顺序恰为 1,2,3,4 的概率是多少？

8. 一口袋中装有 s 个白球，r 个红球，今随机地取出一球，记下它的颜色，再将其放回，然后，再随机地取出一个球，并记下它的颜色，……, 如此下去，直到记录中出现 k 个白球为止，求必须的抽取次数为 T 的概率 (所谓 “随机地” 是指 “等可能随机地”).

9. 设有 n 个房间，分给 n 个不同的人，每个人都以 $\frac{1}{n}$ 的概率进入每一房间，而且每间房里的人数没有限制，试求不出现空房的概率.

10. 将 n 个球投入 N 个匣中，球可辨 (就是说，不仅区别每个匣子中的球数，而且还区别是那几个球), 每个球都以 $1/N$ 的概率进入每一个匣内，求指定某匣是空的概率.

11. 在线段 $(0,a)$ 上任意投三个点，试求由 0 到三点的三线段能构成三角形的概率.

12. 在整数 (0 至 9) 中，任取四个，能排成一个四位偶数的概率是多少？

13. 从数 $1,2,\cdots,N$ 中随机的取出 n 个 (每次取完以后都放回), 问 (i) 所取的 n 个数全不相同的概率是多少？ $(N\geqslant n)$

(ii) 若将取出的数从小到大排列，第 m 个数等于 M 的概率是多少？

14. 将 n 个球投入 N 个匣子，若球不可辨 (就是说，只记每匣中的球数，而不区别是那些球)，每一种投法都是等概率的，试求：

(i) 没有空匣的概率 (这时假设 $n \geqslant N$);

(ii) 正好有 m 个空匣的概率；

(iii) 指定 m 个匣中正好共有 j 个球的概率.

15. 有 k 个坛子，每一个装有 n 个球，分别编有自 1 至 n 的号码，今从每一个坛子中取出一个球，问 m 是所取的球中的最大编号的概率是多少？

16. 从装有 m 个白球与 n 个黑球的坛子中将球一个一个地取出 (每次都放回)，一直取到出现白球为止.

(i) 试写出这一随机试验的基本事件空间，这一随机试验是不是古典型的？

(ii) 求取出的黑球恰好是 k 个的概率.

17. 两船欲靠同一码头. 设二船独立地到达，而且各自到达时间在一昼夜间是等可能的 (即均匀分布的). 如果此二船在码头的停留时间分别是 1 及 2 小时，试求一船要等待空出码头的概率.

18. 两人约好在某一地点会面，在时间 T_1 到 T_2 之中到达. 假定他们在 T_1 至 T_2 之间的任何时刻到达都是等可能的，试求其中一人必须等另外一人的时间不小于 t_0 的概率.

19. $N+1$ 个坛子 $U_0, U_1, \cdots, U_N$, U_i 中含有 i 个白球和 $N-i$ 个黑球，先任选取一个坛子，然后从中取 $n(n<N)$ 个球；若有 v 个白球，$n-v$ 个黑球. 试求

(i) 所取的坛子是 U_i 的概率.

(ii) 再在此坛中取一个球是白球的概率.

20. 有四个口袋，内装白球和黑球，数目如下：(1,2),(2,1),(2,2),(3,1), 今自每个口袋各取一个球，求恰有二个白球的概率.

21. 卜里耶 (Polya) 口袋问题：有些人把这个问题当作传染病或地震的模型，认为某地越爆发则越容易爆发. 这个模型如下 (红球代表爆发地震，黑球表不爆发): 口袋里装有 b 个黑球，r 个红球，任意取出一个，然后放回并再放入 c 个与取出的颜色相同的球，再向袋里取出一球. 问

(i) 最初取出的球是黑的，第二次取出的也是黑色的概率.

(ii) 如将上述手续进行 n 次，取出的正好是 n_1 个黑球，n_2 个红球 $(n_1+n_2=n)$ 的概率.

22. 问题同上，证明：任何一次取得黑球的概率都是 $\frac{b}{b+r}$.

23. 问题同上，试证

(i) 第 m 次与第 n 次 $(m<n)$ 取出都是黑球的概率是 $\frac{b(b+c)}{(b+r)(b+r+c)}$.

(ii) 第 m 次第 n 次取出是 (红，黑) 的概率是 $\frac{br}{(b+r)(b+r+c)}$.

24. 问题同上，设前 n 次抽球中出现 k 个黑球的概率是 $p_k(n)$, 证明递推公式

$$p_k(n+1)=p_k(n)\frac{r+(n-k)c}{b+r+nc}+p_{k-1}(n)\frac{b+(k-1)c}{b+r+nc}.$$

25. 今有两名射手轮流对同一目标进行射击，甲射手命中的概率为 P_1, 乙射手命中的概率为 P_2. 甲先射，谁先命中谁得胜. 问甲，乙得胜的概率各为多少？

26. 口袋里装有 a 个白球及 b 个黑球，任意取出 k 个球，不看取出的球，再从剩下的球中取出一个来，试证取出的是白球的概率为 $\frac{a}{a+b}$, 与 k 无关 (由此可见：第一人抽取白球的概率为 $\frac{a}{a+b}$, 取后不还原；第二人不看第一人取出的结果，再取出白球的概率仍为 $\frac{a}{a+b}$; $\cdots$, 对于第 k 人概率还是 $\frac{a}{a+b}$).

27. 口袋里有 a 个白球 b 个黑球，一个一个地将球取出，不放回，一直取到袋中只剩下相同颜色的球为止. 试求：袋中剩的是白球的概率.

28. 利用概率论的想法证明下列恒等式：设 $A > a$

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots+\frac{(A-a)\cdots 2\cdot 1}{(A-1)\cdots(a+1)a}=\frac{A}{a}.$$

29. 一工人看管三台机床，在一小时内机床不需要工人照顾的概率对于第一台是 0.9, 第二台是 0.8, 第三台是 0.85. 在一小时的过程中，试求：(i) 没有一台机床需要照顾的概率；(ii) 至少有一台不需要照顾的概率.

30. 在一个罐子里有 n 张带有号码 1 至 n 的票，一张一张地取出来 (不放回), 求至少有一次所取的票的号码与次数一致的概率.

31. 坛 I, 装有 5 个白球，5 个黑球，从坛 I 中取 5 个球放入空坛 II 中，再从 II 中取 3 个球放入空坛 III, 最后从坛 III 中取出的一个球是白球，试求第一次取出的 5 个球全是白球的概率.

32. 甲和乙比赛射击，每进行一次，胜者得 1 分. 在一次射击中，甲胜的概率是 α, 乙是 β. 设 $\alpha > \beta, (\alpha+\beta=1)$. 射击进行到有一人超过对方 2 分就停止，多得 2 分者为胜. 试分别求甲获胜的概率和乙获胜的概率.

第二章　随机变数与它的分布

§2.1　随　机　变　数

(一) 随机变数的直观背景．在许多实际问题中，常常需要考虑定义在基本事件空间 $\Omega=(\omega)$ 上的函数 $\xi(\omega)$.

例 1　某人自甲地往乙地，可以“乘船”(ω_1) 去，也可以“乘车”(ω_2) 或“乘飞机”(ω_3) 去，如乘船，旅费为 50 元，如乘车或飞机，旅费分别为 100 元和 200 元．这里 $\Omega=(\omega_1,\omega_2,\omega_3)$, 以 ξ 表所需旅费，那么 ξ 可以看成为定义在 Ω 上的函数；如果出现 ω_1, 就有 $\xi=50$, 这可简记为 $\xi(\omega_1)=50$; 类似考虑 ω_2,ω_3. 这样便得到函数 $\xi(\omega)$:

$$\xi(\omega)=\begin{cases}50, & \text{如}\omega=\omega_1\\100, & \text{如}\omega=\omega_2\\200, & \text{如}\omega=\omega_3\end{cases}\tag{1}$$

如果已知乘船，乘车，乘飞机的概率分别为 $\frac{1}{2},\frac{1}{3},\frac{1}{6}$, 那么旅费为 50 的概率等于 $\frac{1}{2}$, 即

$$P(\xi=50)=\frac{1}{2}.\tag{2}$$

同样，

$$P(\xi=100)=\frac{1}{3},\quad P(\xi=200)=\frac{1}{6}.\tag{3}$$

例 2　从某学校任选一人 ω 而登记他的身长 $\xi(\omega)$, 那么 $\xi(\omega)$ 随着 ω 而变，故是 ω 的函数；同样，体重 $\eta(\omega)$ 也是 ω 的函数.

这种函数不胜枚举，长江的年流量，某地区的一年的雨量，室内的温度，射击的偏差度等等都是.

一般地，设 E 为任一随机试验，它的基本事件空间为 $\Omega=(\omega)$, 如果对每 $\omega\in\Omega$, 有一实数 $\xi(\omega)$ 和它对应，我们就得到一个定义在 Ω 上的实值函数 $\xi(\omega)$. 在概率论中，我们不仅关心 $\xi(\omega)$ 取什么数为值，而且还关心它取某些数为值的可能性的大小，也就是说，还关心它以多少概率取某些数为值 (参看 (2) 式). 例如，我们希望知道 $\omega-$ 集 $(\omega:\xi(\omega)\leqslant x)$ 的概率，其中 x 是任一实数 (以后 $(\omega:\xi(\omega)\leqslant x)$,$(\omega:\xi(\omega)\in A)$ 等等恒简记为 $(\xi(\omega)\leqslant x)$,$(\xi(\omega)\in A)$ 等等). 这一要求带来了一定的数学上的困难.

问题在于：为了要求出 $(\xi(\omega)\leqslant x)$ 的概率，先决条件是 $P(\xi(\omega)\leqslant x)$ 必须有意义，但如 §1.3 所述，我们只是对 Ω 的某 $\sigma-$ 代数 $\mathcal{F}$ 中的 $\omega-$ 集 — 即事件 — 定义了概率 P, 因此先决条件化为 $(\xi(\omega)\leqslant x)\in\mathcal{F}$, 然而，对任意的函数 $\xi(\omega)$ 这个条件未必满足，于是有必要在定义中引进这一条件.

上面直观的想法引出下列正式的定义：

定义 1　设 $(\Omega,\mathcal{F},P)$ 是一概率空间，$\xi(\omega)$ 是定义在 $\Omega=(\omega)$ 上的单值实函数，如果对任一实数 x, $\omega-$ 集 $(\xi(\omega)\leqslant x)$ 是一事件，亦即

$$(\xi(\omega)\leqslant x)\in\mathcal{F},\tag{4}$$

就称 $\xi(\omega)$ 为随机变数.

我们指出，随机变数 $\xi(\omega)$ 并不是自变数，它是满足条件 (4) 的 ω 函数，自变数是 ω.

以后没有必要强调 ω 时，常省去 ω, 而记 $\xi(\omega)$ 为 ξ, $(\xi(\omega)\leqslant x)$ 为 $(\xi\leqslant x)$ 等.

注意 1 由 (4) 可见，随机变数的概念是对于已给的 $\sigma-$ 代数 $\mathcal{F}$ 而言的. 现在考虑一特殊情况：设 $\mathcal{F}$ 由 Ω 中全体子集构成，显然 $\mathcal{F}$ 也是 Ω 中的 $\sigma-$ 代数，这时由于 Ω 的任一子集都是一事件，故任一单值实函数 $\xi(\omega)$ 都满足 (4), 从而是一随机变数.

注意 2 特别，对于只含有穷或可列多个点的 Ω, 我们以后总设 $\mathcal{F}$ 是由 Ω 中的全体子集所成的 $\sigma-$ 代数，因而可利用注意 1 的结果，这一点以后常用.

随机变数概念的产生是概率论发展史中的重大事件，它使概率论研究的对象由事件扩大为随机变数. 设 A 为任一事件，因而 $A\in\mathcal{F}$, 它对应于一个函数 $\chi_A(\omega)$:

$$\chi_A(\omega)=\begin{cases}1, & \text{如}\omega\in A;\\ 0, & \text{如}\omega\in\overline{A}.\end{cases}\tag{5}$$

通常称 $\chi_A(\omega)$ 为 A 的示性函数. 由于

$$(\chi_A(\omega)\leqslant x)=\begin{cases}\Omega\in\mathcal{F}, & \text{如}x\geqslant 1;\\ \overline{A}\in\mathcal{F}, & \text{如}1>x\geqslant 0;\\ \phi\in\mathcal{F}, & \text{如}0>x.\end{cases}$$

可见 $\chi_A(\omega)$ 是一随机变数. 等式

$$P(A)=P(\chi_A(\omega)=1)$$

表示：求 A 的概率 $P(A)$ 等价于求随机变数 $\chi_A(\omega)$ 等于 1 的概率. 这样一来，事件的研究就可纳入随机变数的研究之中. 从现在起，我们的 **研究对象主要集中在随机变数及它的分布上**.

(二) **随机变数的分布函数**. 设 $\xi(\omega)$ 是随机变数，由 (4), $P(\xi(\omega)\leqslant x)$ 有意义，因而可定义函数

$$F_\xi(x)=P(\xi(\omega)\leqslant x),\quad (x\in R_1=(-\infty,\infty))\tag{6}$$

并称 $F_\xi(x)$ 为 ξ 的分布函数. $F_\xi(x)$ 给出了 ξ 取不大于 x 值的概率. 以下会看到，通过 $F_\xi(x)$ 还可表达许多有关 ξ 的事件的概率，例如 $P(\xi(\omega)=x)$ 等等.

定理 1 函数 $F_\xi(x)(x\in R_1)$ 具有下列性质：

(i) 单调不减：如 $b>a$, 则 $F_\xi(b)\geqslant F_\xi(a)$.

(ii) 右连续 [1]: $\lim\limits_{b\to a+0}F_\xi(b)=F_\xi(a)$.

(iii) 令 $F_\xi(-\infty)=\lim\limits_{x\to-\infty}F_\xi(x)$, $F_\xi(\infty)=\lim\limits_{x\to\infty}F_\xi(x)$, 则

$$F_\xi(-\infty)=0,\ F_\xi(\infty)=1.\tag{7}$$

证 (i) 如 $b>a$, 则

$$(\xi(\omega)\leqslant b)=(\xi(\omega)\leqslant a)\cup(a<\xi(\omega)\leqslant b)\tag{8}$$

[1] $b\to a+0$, 表示 $b>a$, 同时 $b\to a$; 类似地， $b\to a-0$ 表 $b<a$, 同时 $b\to a$

右方的二 ω- 集不相交，故

$$F_\xi(b) = P(\xi(\omega) \leqslant b) = P(\xi(\omega) \leqslant a) + P(a < \xi(\omega) \leqslant b)$$
$$\geqslant P(\xi(\omega) \leqslant a) = F_\xi(a).$$

(ii) 由于 $F_\xi(x)$ 的单调性，存在右极限

$$F_\xi(a+0) = \lim_{b \to a+0} F_\xi(b).$$

为了证明 $F_\xi(a+0) = F_\xi(a)$, 根据单调性，只须对某一列 $\{b_n\}$, $b_1 > b_2 > \cdots$, $b_n \to a$, 有 $\lim\limits_{n\to\infty} F_\xi(b_n) = F_\xi(a)$ 即可. 令

$$A_n = (a < \xi(\omega) \leqslant b_n),$$

显然 $A_n \supset A_{n+1}$, $\bigcap\limits_{n=1}^{\infty} A_n = \phi$, 于是由 (8) 及 §1.3 连续性定理得

$$\lim_{n\to\infty} F_\xi(b_n) - F_\xi(a) = \lim_{n\to\infty} [F_\xi(b_n) - F_\xi(a)]$$
$$= \lim_{n\to\infty} P(A_n) = P(\phi) = 0.$$

(iii) 仍然利用单调性：为证 $F_\xi(-\infty) = 0$, 只要证 $\lim\limits_{n\to\infty} F_\xi(-n) = 0$. 考虑 $A_n = (\xi(\omega) \leqslant -n)$, 则 $A_n \supset A_{n+1}$, $\bigcap\limits_{n=1}^{\infty} A_n = \phi$, 由连续性定理

$$\lim_{n\to\infty} F_\xi(-n) = \lim_{n\to\infty} P(A_n) = P(\phi) = 0.$$

$F_\xi(\infty) = 1$ 的证明类似，只要以 §1.3 系 3 代替连续性定理. □

还可以用 $F_\xi(x)$ 来表达一些重要事件的概率：

(a) $P(a < \xi(\omega) \leqslant b) = F_\xi(b) - F_\xi(a)$. (9)

(b) 令 $F_\xi(b-0) = \lim\limits_{a \uparrow b} F_\xi(a)$, 则 $P(\xi(\omega) = b) = F_\xi(b) - F_\xi(b-0)$. (10)

(c) $P(\xi(\omega) < b) = F_\xi(b-0)$. (11)

(d) $P(\xi(\omega) > b) = 1 - F_\xi(b)$. (12)

(e) $P(\xi(\omega) \geqslant b) = 1 - F_\xi(b-0)$. (13)

实际上： (9) 由 (8) 推出. 任取一列单调上升而趋于 b 的数列 a_n, 有

$$(\xi(\omega) = b) = \bigcap_{n=1}^{\infty} (a_n < \xi(\omega) \leqslant b). \tag{14}$$

利用连续性定理及 (9) 得

$$P(\xi(\omega) = b) = \lim_{n\to\infty} P(a_n < \xi(\omega) \leqslant b) = \lim_{n\to\infty} [F_\xi(b) - F_\xi(a_n)].$$

由此得证 (10). 又因 $(\xi(\omega) = b) \subset (\xi(\omega) \leqslant b)$,

$$(\xi(\omega) < b) = (\xi(\omega) \leqslant b) - (\xi(\omega) = b).$$

由此及 (10) 得

$$P(\xi(\omega)<b)=F_\xi(b)-[F_\xi(b)-F_\xi(b-0)]=F_\xi(b-0).$$

最后，由

$$(\xi(\omega)>b)=\Omega-(\xi(\omega)\leqslant b),\ (\xi(\omega)\geqslant b)=\Omega-(\xi(\omega)<b),$$

并利用 $P(\Omega)=1$ 及 (11), 得证 (12)(13).

把 (6) 式改写为

$$F_\xi(x)=P(\xi(\omega)\in(-\infty,x]).$$

这说明作为点 $x\in R_1$ 的函数 $F_\xi(x)$ 可以看成区间 $(-\infty,x]$ 的函数，下面看到，甚至对任一波雷耳集 (定义见下)A, $(\xi(\omega)\in A)$ 都有定义，这样便产生随机变数 ξ 的分布的概念. 为此需要作一些准备工作.

(三) **预备知识**. 先引进一些概念和引理.

设 $\Omega=(\omega)$ 是任意给定的空间， $\mathcal{G}$ 是 Ω 的一个子集系，也就是说， $\mathcal{G}$ 是 Ω 的某些子集所成的集. 则必至少存在 Ω 中的一个 $\sigma-$ 代数 $\mathfrak{M}$, 使 [1] $\mathfrak{M}\supset\mathcal{G}$. 实际上，只要取 $\mathfrak{M}$ 为 Ω 的全体子集所成的 $\sigma-$ 代数 (它是最大的 $\sigma-$ 代数) 就行了.

以 $\sigma(\mathcal{G})$ 表一切含 $\mathcal{G}$ 的 $\sigma-$ 代数 $\mathfrak{M}$ 的交 [2] , 即:

$$\sigma(\mathcal{G})=\bigcap_{\mathfrak{M}\supset\mathcal{G}}\mathfrak{M}\tag{15}$$

引理 1 $\sigma(\mathcal{G})$ 是 Ω 中的 $\sigma-$ 代数.

证 由于 Ω 属于每一 $\mathfrak{M}$, 故 $\Omega\in\sigma(\mathcal{G})$. 设 $A\in\sigma(\mathcal{G})$, 则 $A\in\mathfrak{M}$. 既然 $\mathfrak{M}$ 是 $\sigma-$ 代数，故 $\overline{A}\in\mathfrak{M}$, 这对一切含 $\mathcal{G}$ 的 $\mathfrak{M}$ 成立，故 $\overline{A}\in\bigcap\limits_{\mathfrak{M}\supset\mathcal{G}}\mathfrak{M}=\sigma(\mathcal{G})$. 最后，如 $A_i\in\sigma(\mathcal{G})$, 则 $A_i\in\mathfrak{M}$, $i=1,2,\cdots$, 与上同理 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathfrak{M}$ 对一切含 $\mathcal{G}$ 的 $\mathfrak{M}$ 成立，故 $\bigcup\limits_{i=1}^{\infty}A_i\in\bigcap\limits_{\mathfrak{M}\supset\mathcal{G}}\mathfrak{M}=\sigma(\mathcal{G})$. □

如果 $\sigma-$ 代数 $\mathfrak{M}(\mathcal{G})$ 具有下二性质，就称它为含 $\mathcal{G}$ 的最小 $\sigma-$ 代数. 1° $\mathfrak{M}(\mathcal{G})\supset\mathcal{G}$; 2° 如 $\mathfrak{M}$ 是任一含 $\mathcal{G}$ 的 $\sigma-$ 代数，则 $\mathfrak{M}\supset\mathfrak{M}(\mathcal{G})$.

引理 2 $\mathfrak{M}(\mathcal{G})=\sigma(\mathcal{G})$; 这就是说：含 $\mathcal{G}$ 的最小 $\sigma-$ 代数重合于一切含 $\mathcal{G}$ 的 $\sigma-$ 代数的交.

证 由于 $\mathfrak{M}(\mathcal{G})\supset\mathcal{G}$, 故由 $\sigma(\mathcal{G})$ 的定义知 $\mathfrak{M}(\mathcal{G})\supset\sigma(\mathcal{G})$. 反之，由引理 1 知 $\sigma(\mathcal{G})$ 是 $\sigma-$ 代数，由 (15) 知 $\sigma(\mathcal{G})\supset\mathcal{G}$, 从而由 2° 知 $\sigma(\mathcal{G})\supset\mathfrak{M}(\mathcal{G})$. 综合两方面即得所欲证. □

特别，如 $\mathcal{G}$ 是 $\sigma-$ 代数，则显然 $\mathfrak{M}(\mathcal{G})=\sigma(\mathcal{G})=\mathcal{G}$.

考虑一重要的特殊情形：设 $\Omega=R_n$, R_n 表 n 维空间，这时每一 $\omega=(x_1,\cdots,x_n)$ 是 n 维空间中的点， x_i 是实数， $i=1,\cdots,n$. 对我们重要是的所谓 n 维波雷耳 (Borel) 集，现在来叙述它们的定义，先从一维情况开始.

设 $\Omega=R_1$. 考虑 R_1 中右半闭的无穷区间 $(-\infty,a]$, 全体这样的区间构成 R_1 中的子集系 $\mathcal{G}$, 即

$$\mathcal{G}=\{(-\infty,a],\ a\in R_1\}\tag{16}$$

---

[1]二集系 $G,\mathcal{G}$ 有关系 $G\supset\mathcal{G}$ 是说：如集 $A\in\mathcal{G}$, 则 $A\in G$

[2]二集系 $\mathfrak{M}_1,\mathfrak{M}_2$ 的交定义为全体满足 $A\in\mathfrak{M}_1$ 及 $A\in\mathfrak{M}_2$ 的集 A 所成的集系，记为 $\mathfrak{M}_1\cap\mathfrak{M}_2$. 类似定义任意多个集系的交

称 $\sigma(\mathcal{G})$ 为波雷耳 $\sigma-$ 代数. 通常记为 $\mathcal{B}_1$.$\mathcal{B}_1$ 中的集称为波雷耳集 [1].

设 $f(x)$ 是定义在 R_1 上的单值函数，取实数值，如果对任意实数 a, 有

$$(f(x) \leqslant a) \in \mathcal{B}_1 \tag{17}$$

则称 $f(x)$ 为波雷耳可测函数，或 $\mathcal{B}_1-$ 可测函数 [2].

今设 $\Omega = R_n$, n 为任一正整数，R_n 中的右半闭无穷区间记为 $\prod\limits_{i=1}^{n}(-\infty, a_i]$, 它是一 n 维点集，由如下的点 $(x_1, \cdots, x_n)$ 构成；$-\infty < x_i \leqslant a_i$, $i = 1, \cdots, n$. 全体这样的区间构成 R_n 中的子集系 $\mathcal{G}$:

$$\mathcal{G} = \left\{ \prod_{i=1}^{n}(-\infty, a_i],\ a_i \in R_1, i = 1, \cdots, n \right\} \tag{18}$$

称 $\sigma(\mathcal{G})$ 为 n 维波雷耳 $\sigma-$ 代数，通常记为 $\mathcal{B}_n$, $\mathcal{B}_n$ 中的集称为 n 维波雷耳集；当 $n = 1$ 时，常将“一维”二字省去.

设 $f(x_1, \cdots, x_n)$ 是定义在 R_n 上的单值函数，取实数值，如果对任意实数 a, 有

$$(f(x_1, \cdots, x_n) \leqslant a) \in \mathcal{B}_n \tag{19}$$

则称 $f(x_1, \cdots, x_n)$ 为 n 元波雷耳可测函数，或 $\mathcal{B}_n-$ 可测函数.

(四) 随机变数的分布. 继续考虑定义在 $(\Omega, \mathcal{F}, P)$ 上的随机变数 $\xi(\omega)$.

引理 3 为使 (4) 式对任意 $x \in R_1$ 成立，充要条件是对任意 $A \in \mathcal{B}_1$, 有

$$(\xi(\omega) \in A) \in \mathcal{F} \tag{20}$$

证 充分性：设 (20) 对任一 $A \in \mathcal{B}_1$ 成立，在 (20) 中取 $A = (-\infty, x]$, 即得 (4).

必要性：令

$$\mathfrak{M} = \{A:\ A \subset R_1, (\xi(\omega) \in A) \in \mathcal{F}\}, \tag{21}$$

换句话说，$\mathfrak{M}$ 是由 R_1 中一切使 (20) 式成立的集 A 构成的集系，试证 $\mathfrak{M}$ 是 R_1 中的 $\sigma-$ 代数. 实际上：$(\xi(\omega) \in R_1) = \Omega \in \mathcal{F}$, 故 $R_1 \in \mathfrak{M}$; 其次，如 $A \in \mathfrak{M}$, 于是 A 满足 (21) 右方括号中二条件，故 $\overline{A} = R_1 - A \subset R_1$; 再注意 $\mathcal{F}$ 是 $\sigma-$ 代数，得 $(\xi(\omega) \in \overline{A}) = \overline{(\xi(\omega) \in A)} \in \mathcal{F}$, 因此 $\overline{A}$ 也满足此二条件，从而 $\overline{A} \in \mathfrak{M}$; 第三，设 $A_i \in \mathfrak{M}$, $i = 1, 2, \cdots$, 显然 $\bigcup\limits_{i=1}^{\infty} A_i \subset R_1$, $\left(\xi(\omega) \in \bigcup\limits_{i=1}^{\infty} A_i\right) = \bigcup\limits_{i=1}^{\infty} (\xi(\omega) \in A_i) \in \mathcal{F}$, 故 $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{M}$.

再证 $\mathfrak{M}$ 包含 R_1 中全体区间 $(-\infty, x]$: 由 (4)

$$(\xi(\omega) \in (-\infty, x]) = (\xi(\omega) \leqslant x) \in \mathcal{F},$$

故 $(-\infty, x]$ 满足 (21) 右方二条件，于是 $(-\infty, x] \in \mathfrak{M}$.

这样便证明了：$\mathfrak{M}$ 是包含 $\mathcal{G} = \{(-\infty, x], x \in R_1\}$ 的 $\sigma-$ 代数. 于是由 (15) 即知

$$\mathfrak{M} \supset \sigma(\mathcal{G}) = \mathcal{B}_1;$$

---

[1] 不难证明，这里波雷耳集的定义与实变函数论中的定义是一致的. 特别，一切开集，闭集，以及开集或闭集的可列次和集 (或可列次交集) 都是波雷耳集

[2] 一切连续函数，连续函数列的极限函数都是 $\mathcal{B}_1-$ 可测函数

这就是说，任一 $A\in\mathcal{B}_1$ 必满足 (21) 右方二条件；从而 (20) 式对任意 $A\in\mathcal{B}_1$ 成立. □

引理 3 的证明方法是典型的，值得注意.

今定义集函数

$$F_\xi(A)=P(\xi(\omega)\in A),\quad A\in\mathcal{B}_1. \tag{22}$$

由于 (20),$F_\xi(A)$ 对任意 $A\in\mathcal{B}_1$ 有意义，我们称此集函数 $F_\xi(A)\,(A\in\mathcal{B}_1)$ 为随机变数 ξ 的分布，它有如下性质：

定理 2 集函数 $F_\xi(A)$ 是 $\mathcal{B}_1$ 上的概率测度.

证 由 (22) 显然可见 $0\leqslant F_\xi(A)\leqslant 1$;

$$F_\xi(R_1)=P(\xi(\omega)\in R_1)=P(\Omega)=1;$$

如 $A_i\in\mathcal{B}_1, i=1,2,\cdots,A_iA_j=\phi,i\neq j$, 则

$$F_\xi\left(\bigcup_{i=1}^{\infty}A_i\right)=P\left(\xi(\omega)\in\bigcup_{i=1}^{\infty}A_i\right)=P\left(\bigcup_{i=1}^{\infty}(\xi(\omega)\in A_i)\right)$$
$$=\sum_{i=1}^{\infty}P(\xi(\omega)\in A_i)=\sum_{i=1}^{\infty}F_\xi(A_i).\quad\square$$

§2.2 分布与分布函数

(一) **分布与分布函数的关系**. 在 §2.1 中，我们从已给的随机变数 $\xi(\omega)$ 出发，定义了 $\xi(\omega)$ 的分布与 $\xi(\omega)$ 的分布函数，其实“分布”与“分布函数”的概念也可不依赖于 $\xi(\omega)$ 而单独引进.

考虑一维空间 R_1 及其中的波雷耳 $\sigma-$ 代数 $\mathcal{B}_1$, 定义在 $\mathcal{B}_1$ 上的概率测度 $F(A)\ (A\in\mathcal{B}_1)$ 称为一维分布，简称分布；定义在 R_1 上的函数 $F(x)$, 如果它具有 §2.1 定理 1 中三性质 (i)(ii)(iii), 即单调不减，右连续而且 $F(-\infty)=0$, $F(\infty)=1$, 就称为一元分布函数或简称为分布函数.

由 §2.1 定理 1,2 可见，随机变数 $\xi(\omega)$ 的分布 $F_\xi(A)$ 是一分布， $\xi(\omega)$ 的分布函数 $F_\xi(x)$ 是一分布函数.

定理 1 分布与分布函数是一一对应的.

***证** 设已给分布为 $F(A), A\in\mathcal{B}_1$, 定义函数

$$F(x)=F((-\infty,x]),\ (x\in R_1) \tag{1}$$

由于 $F(A)\ (A\in\mathcal{B}_1)$ 是一概率测度，依照 §2.1 定理 1 的证明 [1] , 易见 $F(x)$ 是一分布函数.

反之，设已给一分布函数 $F(x),(x\in R_1)$, 以 $F(A)$ 表由 $F(x)$ 所产生的勒贝格 - 司梯阶 (Lebesgue-Stieltjes) 测度 [2]. $F(A)$ 对任一波雷耳集 A 有定义，于是 $F(A),(A\in\mathcal{B}_1)$ 是 $\mathcal{B}_1$ 上的测度，它满足 (1). 由于

$$F(R_1)=\lim_{x\to\infty}F((-\infty,x])=\lim_{x\to\infty}F(x)=1$$

---

[1] 以 $(-\infty,b],(a,b],F((-\infty,x])$ 分别代替那里的 $(\xi(\omega)\leqslant b),(a<\xi(\omega)\leqslant b),P(\xi(\omega)\leqslant x)$ 等等.

[2] 参看 [10] 第五章 §9, 关于勒贝格 - 司梯阶测度 (和以后用到的勒贝格 - 司梯阶积分) 的更详细的介绍见 [3]. 这测度定义在包括全体波雷耳集在内的 $L-S$ 可测 (与 $F(x)$ 有关) 集合类上.

可见 $F(A)$ 还是 $\mathcal{B}_1$ 上的概率测度. 最后注意：由勒贝格 - 司梯阶测度的理论，知对已给的 $F(x)$, 满足 (1) 的概率测度 $F(A)(A \in \mathcal{B}_1)$ 是唯一的. □

我们知道，随机变数 $\xi(\omega)$ 的分布函数 $F_\xi(x)$ 是分布函数，反之，下面的结论也成立.

***定理 2** (存在定理) 设已给分布函数 $F(x)$, $(x \in R_1)$, 则必存在概率空间 $(\Omega, \mathcal{F}, P)$ 及定义于其上的随机变数 $\xi(\omega)$, $(\omega \in \Omega)$ 使 $\xi(\omega)$ 的分布函数 $F_\xi(x)$ 重合于 $F(x)$, 即

$$F_\xi(x) = F(x) \quad (\text{一切} x \in R_1). \tag{2}$$

证 首先应造 $(\Omega, \mathcal{F}, P)$ 及 $\xi(\omega)$. 令 $\Omega = R_1$, $\mathcal{F} = \mathcal{B}_1$, P 为由已给的分布函数 $F(x)$ 所产生的勒贝格 - 司梯阶测度 $F(A)$ 在 $\mathcal{B}_1$ 上的限制 [1], 即 $P(A) = F(A)$, $(A \in \mathcal{B}_1)$, 因而

$$F((-\infty, x]) = F(x). \tag{3}$$

由于 $\Omega = R_1$, 对 Ω 中的点 $\omega \in R_1$, 定义

$$\xi(\omega) = \omega. \tag{4}$$

我们证明：函数 $\xi(\omega) = \omega$ 是所求的随机变数. 首先注意，对任意 $x \in R_1$,

$$(\xi(\omega) \leqslant x) = (\omega \leqslant x) = (-\infty, x] \in \mathcal{B}_1,$$

故 $\xi(\omega)$ 是一随机变数；其次，由上式及 (3) 得

$$F_\xi(x) = P(\xi(\omega) \leqslant x) = P((-\infty, x]) = F((-\infty, x]) = F(x),$$

故 (2) 式也成立. □

考虑分布函数 $F(x)$ 及任一点 $a \in R_1$, 令

$$C(a) = F(a) - F(a-0), \tag{5}$$

称此数 $C(a)$ 为 $F(x)$ 在点 a 的跃度. 由 $F(x)$ 的单调不减性， $C(a) \geqslant 0$. 如果 $C(A) > 0$, 则称 a 为 $F(x)$ 的跳跃点.

***定理 3** 分布函数 $F(x)$ 的不连续点所成的集是有限集或可列集.

证 由分布函数的单调不减性， $F(x)$ 的不连续点集 D 重合于跳跃点集，因为 $0 \leqslant F(x) \leqslant 1$, 可见跃度不小于 $\frac{1}{2^n}$ 的跳跃点所成的集 D_n 至多只含 2^n 个点，然而 $D = \bigcup_{n=1}^{\infty} D_n$, 可见 D 是有限集或可列集. □

由 (1) 看出：可把分布函数看成区间 $(-\infty, x]$ 的函数，当分布 $F(A)$ 的定义域缩小到形如 $(-\infty, x]$ 的集时，便得到分布函数，故分布的定义域更大，因此，分布是比分布函数更精细的概念. 然而在运算中人们往往更多地用分布函数，原因在于，一般说来点函数比集函数容易掌握.

为了便于理解分布与分布函数二概念的实质，我们给出一个直观的解释：在全直线 $R_1 = (-\infty, \infty)$ 上，用任意一种散布方式布上质量为 1 的物质 (例如 1 克面粉), 以 $F(A)$ 表

[1] $F(A)$ 的定义域包含 $\mathcal{B}_1$, 当我们只在 $\mathcal{B}_1$ 上考虑 $F(A)$ 时，它是 $\mathcal{B}_1$ 上的测度，称此测度为 $F(A)$ 在 $\mathcal{B}_1$ 上的限制 (测度)

分布在波雷耳集 A 中的物质的质量，那么 $F(A)(A\in\mathcal{B}_1)$ 就成一分布；以 $F(x)$ 表 $(-\infty,x]$ 中的质量，$F(x)(x\in R_1)$ 是一分布函数．反之，设已给 $\mathcal{B}_1$ 上一分布 $F(A)$(或 R_1 上一分布函数 $F(x)$), 将质量为 1 的物质散布在 R_1 上，使任一波雷耳集 A(或任一 $(-\infty,x]$) 所含物质质量恰为 $F(A)$(或恰为 $F(x)$), 这样便决定了一种散布方式．由此可见：分布 (或分布函数) 与散布方式是一一对应的．这也正是分布 (与分布函数) 命名的根据．

(二) **离散型分布**．实际中常遇到的两种重要的分布函数 (或分布) 是离散型的与连续型的．

先考虑一简单的例：令

$$F(x)=\sum_{0<n\leqslant x}\frac{1}{2^n},$$

其中求和对一切不大于 x 的正整数进行．显然 $F(x)$ 单调不减，右连续，$F(-\infty)=0$, $F(\infty)=1$, 故它是一分布函数．

一般地，称分布函数 $F(x)(x\in R_1)$ 为离散型的，如果存在一个 (横) 行数为 2,(直) 列数为有穷或可列的矩阵

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}, \tag{6}$$

其中

$$a_i\in R_1,\ P_i\geqslant 0,\ \sum_{i=0}^{\infty}p_i=1, \tag{7}$$

使对任意 $x\in R_1$, 有

$$F(x)=\sum_{i:a_i\leqslant x}p_i. \tag{8}$$

这里求和对一切满足 $a_i\leqslant x$ 的 i 进行 (如果这样的 i 不存在，就令 $F(x)=0$). 例如在上例中，

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}=\begin{pmatrix} 1 & 2 & 3 & \cdots \\ \frac{1}{2} & \frac{1}{2^2} & \frac{1}{2^3} & \cdots \end{pmatrix}.$$

称分布 $F(A)(A\in\mathcal{B}_1)$(或随机变数 ξ) 是离散型的，如果它的分布函数是离散型的．

称矩阵 (6) 为此分布函数 (或分布，或随机变数) 的密度矩阵，简称密度．

以后用 $\{a\}$ 表仅由一点 a 构成的单点集，记

$$\Lambda=(a_0,a_1,a_2,\cdots).$$

并以 $F(A)(A\in\mathcal{B}_1)$ 表 $F(x)(x\in R_1)$ 对应的分布，由 (8) 得

$$F(\{a_i\})=F(a_i)-F(a_i-0)=p_i, \tag{9}$$

$$F(A)=\sum_{i:a_i\in A}F(\{a_i\})=\sum_{i:a_i\in A}p_i. \tag{10}$$

(10) 中求和对一切满足 $a_i\in A$ 的 i 进行，特别，由 (10) 及 (7)

$$F(\Lambda)=\sum_{i=0}^{\infty}p_i=1. \tag{11}$$

(11) 式表示，按此分布散布的质量集中在有穷或可列集 Λ 上.

一般地，对任一分布 (不论离散与否)$F(A)(A \in \mathcal{B}_1)$, 如存在波雷耳集 B, 使 $F(B)=1$, 就说此分布集中在 B 上.

如果 (6) 是随机变数 ξ 的密度，由 (9) 得

$$P(\xi = a_i) = F(\{a_i\}) = p_i.$$

故 p_i 是 ξ 取值 a_i 的概率，而由 (10) 知 $p(\xi \in A) = \sum\limits_{i:a_i \in A} p_i$.

设 Λ 是有穷集，含点 $a_0, a_1, \cdots, a_n$, 不妨设

$$a_0 < a_1 < \cdots < a_n,\ p_i > 0\ (i = 0, 1, \cdots, n).$$

(有必要时，改变 (6) 中列的次序并删去使 $p_i = 0$ 的列后，上式满足). 这时密度及分布函数的图见图 1, 图 2.

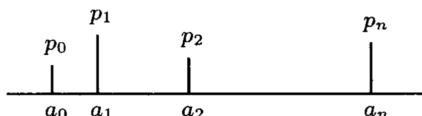


图 1

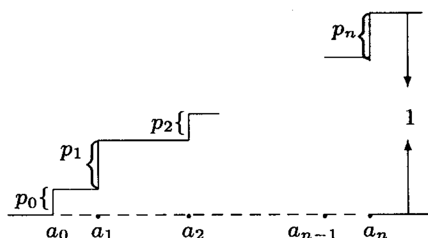


图 2

由此可见，此时 $F(x)$ 是右连续的阶梯函数.

但如 Λ 为可列集时，那么图形可能很复杂，因为 Λ 可以在 R_1 中处处稠密；不过如果在任一半无穷区间 $(-\infty, x]$ 中只含有限多个 a_i 时，则 $F(x)$ 仍是右连续的阶梯函数.

为了给出一离散型分布函数，只要给出一满足 (7) 的矩阵 (6), 然后用 (8) 以定义 $F(x)$, $(x \in R_1)$.

例 1 (单点分布与两点分布) 密度为 $\binom{a_0}{1}$ 的分布叫单点分布，此分布集中在 $\{a_0\}$ 上，分布函数为

$$F(x) = \begin{cases} 0, & 如 x < a_0, \\ 1, & 如 x \geqslant a_0. \end{cases} \tag{12}$$

密度为 $\begin{pmatrix} a_0 & a_1 \\ p & q \end{pmatrix}$ (其中 $p \geqslant 0, q \geqslant 0, p+q=1$)) 的分布叫两点分布，它集中在两点集 $\Lambda = (a_0, a_1)$ 上，如 $a_0 < a_1$, 则分布函数为

$$F(x) = \begin{cases} 0, & 如 x < a_0, \\ p, & 如 a_0 \leqslant x < a_1, \\ 1, & 如 a_1 \leqslant x. \end{cases} \tag{13}$$

(三) 连续型分布. 称分布函数 $F(x)(x \in R_1)$ 为连续型的，如果存在非负的函数 $f(x)(x \in R_1)$, 使对任意实数 x, 有 [1]

[1] 不言而喻，(14) 及下面 (15) 已蕴含 $f(x)$ 在 R_1 上的勒贝格可积性的假定，(14) (15) 中为勒贝格积分. 由 (14)

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \tag{14}$$

称分布 $F(A)(A \in \mathcal{B}_1)$(或随机变数 ξ) 为连续型的，如果它的分布函数是连续型的. 函数 $f(x)(x \in R_1)$ 称为此分布函数 (或分布，或随机变数) 的密度函数，简称密度. 由此可见，连续型分布函数完全为密度所决定.

反之，已给任一非负函数 $f(x)$, 如果

$$\int_{-\infty}^{\infty} f(x)\,dx = 1, \tag{15}$$

那么由 (14) 定义的 $F(x)(x \in R_1)$ 是一连续型的分布函数.

不要把 "连续型的分布函数" 与 "连续的分布函数" 相混淆，后者表示此分布函数 $F(x)$ 对 x 是连续的；前者表示 (14) 式成立，也就是说 $F(x)$ 是绝对连续的. 由此可见：连续型的分布函数必定连续，反之不成立.

例 2 (均匀分布) 设 $-\infty < a < b < \infty$, 令

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & 如 x \in [a,b], \\ 0, & 如\ x \,\overline{\in}\, [a,b]. \end{cases} \tag{16}$$

显然，$f(x)$ 非负而且满足 (15). 以此 $f(x)$ 为密度的分布 $F(A)$, $(A \in \mathcal{B}_1)$ 称为均匀分布，由于

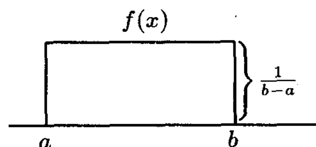


图 3

$$F([a,b]) = \int_a^b \frac{1}{b-a}\,dx = 1,$$

故此分布集中在 $[a,b]$ 上. 它的分布函数为

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \begin{cases} 0, & 如 x < a, \\ \dfrac{x-a}{b-a}, & 如 a \leqslant x \leqslant b, \\ 1, & 如 b < x. \end{cases} \tag{17}$$

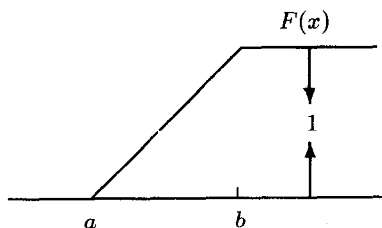


图 4

均匀分布含二参数 $a, b(a < b)$, 当它们确定时，这分布就完全决定了. $f(x)$ 及 $F(x)$ 的图见图 3, 图 4.

例 3 (指数分布) 设 a, b 为二常数，$a > 0$, 令

$$f(x) = \begin{cases} 0, & 如 x \leqslant b, \\ ae^{-a(x-b)}, & 如 x > b. \end{cases} \tag{18}$$

得：几乎处处 (关于勒贝格测度) 有 $F'(x) = f(x)$, 因而 (14) 式等价于

$$F(x) = \int_{-\infty}^{x} F'(y)\,dy.$$

显然， $f(x)$ 非负而且满足 (15). 以此 $f(x)$ 为密度的分布称为指数分布，它含二参数 a,b, $(a>0)$, 它的分布函数为

$$F(x)=\begin{cases}0, & \text{如} x\leqslant b,\\ 1-e^{-a(x-b)}, & \text{如} x>b.\end{cases} \tag{19}$$

离散型与连续型分布虽是两种重要的分布，然而却远未穷尽一切分布，试举一简单的例. 为此先注意下列事实：

设 a_1,a_2 为二常数 $a_1\geqslant 0, a_2\geqslant 0, a_1+a_2=1$; 又设 $F_1(x),F_2(x)$ 为二分布函数，则

$$F(x)=a_1F_1(x)+a_2F_2(x) \tag{20}$$

是一分布函数.

实际上，由 $F_i(x)(i=1,2)$ 的右连续性及单调不减性得知 $F(x)$ 也有此二性质，又

$$\lim_{x\to-\infty}F(x)=a_1\lim_{x\to-\infty}F_1(x)+a_2\lim_{x\to-\infty}F_2(x)=0,$$
$$\lim_{x\to\infty}F(x)=a_1\lim_{x\to\infty}F_1(x)+a_2\lim_{x\to\infty}F_2(x)=a_1+a_2=1.$$

故 $F(x)$ 是分布函数.

例 4 令 $a_1=a_2=\frac{1}{2}$, 又

$$F_1(x)=\begin{cases}0, & \text{如} x<0,\\ 1, & \text{如} x\geqslant 0.\end{cases}$$

$$F_2(x)=\begin{cases}0, & \text{如} x<0,\\ x, & \text{如} 0\leqslant x\leqslant 1,\\ 1, & \text{如} 1<x.\end{cases}$$

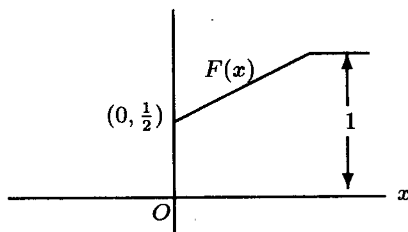


图 5

(分别在 (12) 中令 $a_0=0$ 及在 (16) 中令 $a=0, b=1$ 即得 $F_1(x)$ 及 $F_2(x)$). 这时由 (20) 定义的

$$F(x)=\begin{cases}0, & \text{如} x<0,\\ \dfrac{1+x}{2}, & \text{如} 0\leqslant x\leqslant 1,\\ 1, & \text{如} 1<x.\end{cases}$$

显然， $F(x)$ 所对应的分布不集中在一可列集或有穷集上，故 $F(x)$ 不是离散型的；它也不是连续型的，因为 $F(x)$ 不是连续函数.

*(**四) 奇异型分布**. 理论上很有价值然而实际问题中很少应用的一种分布函数是所谓奇异型的. 称分布函数 $F(x)(x\in R_1)$ 为奇异型的，如果 $F(x)$ 连续，而且它的导函数 [1] 几乎处处 (关于勒贝格测度而言) 等于 0. 称分布 $F(A)$, $(A\in\mathcal{B}_1)$(或随机变数 ξ) 为奇异型的，如果它的分布函数是奇异型的.

著名的奇异型分布函数的例如下：

---

[1]我们知道，单调函数几乎处处可导

例 5 作康脱尔 (Cantor) 函数 $F(x)$, $(x \in R_1)$, 它的定义如下

$$F(x)=\begin{cases} 0, & \text{如} x<0, \\ 1, & \text{如} x>1, \\ \frac{1}{2}, & \text{如} x \in \left(\frac{1}{3}, \frac{2}{3}\right), \\ \frac{1}{4}, & \text{如} x \in \left(\frac{1}{9}, \frac{2}{9}\right), \\ \frac{3}{4}, & \text{如} x \in \left(\frac{7}{9}, \frac{8}{9}\right), \\ \frac{1}{8}, & \text{如} x \in \left(\frac{1}{27}, \frac{2}{27}\right), \\ \frac{3}{8}, & \text{如} x \in \left(\frac{7}{27}, \frac{8}{27}\right), \\ \frac{5}{8}, & \text{如} x \in \left(\frac{19}{27}, \frac{20}{27}\right), \\ \frac{7}{8}, & \text{如} x \in \left(\frac{25}{27}, \frac{26}{27}\right), \\ \cdots\cdots . \end{cases} \tag{21}$$

换句话说, 把 $(0,1)$ 分成三个等长的子区间, 在中间一个子区间 $\left(\frac{1}{3}, \frac{2}{3}\right)$ 上定义 $F(x)$ 为 $\frac{1}{2}(1-0)=\frac{1}{2}$; 把 $\left(0, \frac{1}{3}\right)$, $\left(\frac{2}{3}, 1\right)$ 各分成三个等长的子区间, 在两个中间子区间 $\left(\frac{1}{9}, \frac{2}{9}\right)$ 及 $\left(\frac{7}{9}, \frac{8}{9}\right)$ 上分别定义 $F(x)$ 为 $\frac{1}{2}\left(\frac{1}{2}-0\right)=\frac{1}{4}$ 及 $\frac{1}{2}+\frac{1}{2}\left(1-\frac{1}{2}\right)=\frac{3}{4}$; $\cdots$; 依照这手续定义下去, 结果只在 (21) 右方诸区间的和的补集上未定义. 注意这补集是处处不稠密的完全集, 它的勒贝格测度为 0, 通常称此补集为康脱尔集, 记为 C. 在 C 上按照连续性要求来补定义 $F(x)$, 于是 $F(x)$ 在 R_1 上都有定义, 它的不完全的示意图形如图 6 所示.

显然, $F(x)$ 单调不减, 连续, $F(-\infty)=0$, $F(\infty)=1$, 故是一分布函数. 由于它在 $(-\infty, 0)$, $(1, \infty)$ 上各为常数, 故 $F'(x)$ 在那里等于 0; 同理在 $\left(\frac{1}{3}, \frac{2}{3}\right)$, $\left(\frac{1}{9}, \frac{2}{9}\right)$, $\left(\frac{7}{9}, \frac{8}{9}\right)$, 等区间上 $F'(x)$ 也等于 0. 由于这些区间的总长为 1, 故得证 $F'(x)$ 几乎处处等于 0, 从而 $F(x)$ 是奇异型分布函数.

任一奇异型分布函数 $F(x)$ 不可能是离散型的, 这是因为 $F(x)$ 有连续性, 如果说 $F(x)$



图 6

是连续型的, 即如可表为 (14) 的形状, 那么由 (14) 知几乎处处有

$$f(x)=F'(x)=0.$$

从而 $\int_{-\infty}^{\infty} f(x)\, dx=\int_{-\infty}^{\infty} F'(x)\, dx=0$, 这与 (15) 矛盾, 故 $F(x)$ 也不可能是连续型的. 由此

可见，没有一个分布函数能属于上述三种类型之二，故这三类分布函数是互相排斥的（进一步的讨论见 §2.13).

§2.3　二项分布与贝努里试验

(一) **二项分布**．分布函数是一种特殊的单调函数，因而也是函数论研究的对象，那里对单调函数的一般性质有深入的研究．概率论中也研究分布函数，不过着重点不在于讨论它们的一般性质，而在于挑选一部分在实际中有广泛应用的分布与分布函数，阐明这种分布函数与实际的联系，研究它们的特性，然后运用到实际中去．从这种观点出发，自然应该把分布，分布函数和随机变数，随机试验结合起来考虑，才能达到上述的目的．

我们从二项分布开始，下面看到，产生这种分布的重要现实泉源是所谓贝努里 (Bernoulli) 试验．

设已给三常数：正整数 n, 二非负数 $p,q,p+q=1$. 考虑矩阵

$$\begin{pmatrix} 0 & 1 & \cdots & n \\ p_0 & p_1 & \cdots & p_n \end{pmatrix} \quad 其中 p_k = \binom{n}{k} p^k q^{n-k} \tag{1}$$

显然 $p_k \geqslant 0, k=0,1,\cdots,n$, 而且

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1. \tag{2}$$

因而对此矩阵，§2.2 中 (7) 式满足．

以 (1) 中矩阵为密度矩阵的分布称为二项分布．

这分布依赖于两个参数：正整数 n 及实数 p, $0 \leqslant p \leqslant 1$(注意 $q=1-p$ 由 p 决定). 记此分布为 $B(n,p)$; 为了强调 (1) 中 p_k 对 p,n 的依赖性，记 p_k 为 $p_k(n,p)$, 因而

$$p_k(n,p) = \binom{n}{k} p^k q^{n-k} \quad (k=0,1,\cdots,n). \tag{3}$$

(二) **贝努里试验**．它是产生二项分布的重要现实泉源．

设 E 为一随机试验，只有两个基本事件 A 及 $\overline{A}$. 令 $p=P(A)$, $q=P(\overline{A})$, $0<p<1$, $0<q<1$, $p+q=1$.

将 E 独立地重复 n 次 (或可列次), 称 $\widetilde{E}=E^n$(或 $\widetilde{E}=E^\infty$) 为 n 重 (或可列重) 贝努里试验；没有必要强调重数时，就简称为贝努里试验．例如，§1.5 中的不断射击构成一贝努里试验，这时 E 为一次射击，A 表事件"中的",$\overline{A}$ 表"不中的".

E^n 中基本事件空间 $\widetilde{\Omega}=(\widetilde{\omega})$ 共含 2^n 个点 $\widetilde{\omega}$. $\widetilde{\omega}$ 由若干个 A 和若干个 $\overline{A}$ 构成，它们的总共个数是 n. 例如，E^2 的基本事件空间是

$$\widetilde{\Omega} = (AA, A\overline{A}, \overline{A}A, \overline{A}\,\overline{A}), \tag{4}$$

其中 $A\overline{A}$ 与 $\overline{A}A$ 应看成为两个不同的基本事件，虽然它们有相同的概率为 pq, 因为 $A\overline{A}$ 表示先出现 A, 后出现 $\overline{A}$, 而 $\overline{A}A$ 则相反．

然而 E^{∞} 的基本事件空间 $\widetilde{\Omega}$ 却是不可列的，因为 $\widetilde{\Omega}$ 是一切

$$\widetilde{\omega} = B_1 B_2 \cdots B_i B_{i+1} \cdots, \text{ 其中 } B_i = A \text{ 或 } \overline{A} \tag{5}$$

的集，因而 $\widetilde{\Omega}$ 与区间 $[0,1]$ 是一一对应的 (参看 §1.5 例 3).

伴随着 E^n, 可以定义一个变数 ξ: ξ 表这 n 次试验中， A 出现的总次数，例如当 $n=2$ 时， ξ 定义在 (4) 中的 $\widetilde{\Omega}$ 上：

$$\xi(AA) = 2, \ \ \xi(A\overline{A}) = \xi(\overline{A}A) = 1, \ \ \xi(\overline{A}\,\overline{A}) = 0.$$

根据 §2.1 注意 2, 知 ξ 是随机变数.

现在来求 ξ 的分布，结果发现， ξ 的分布是二项分布 $B(n,p)$, 其中 $p = P(A)$.

注意 $P_k = P(\xi = k)$ $(k = 0,1,\cdots,n)$ 是 n 重贝努里试验 $\widetilde{E} = E^n$ 中，恰好出现 k 次 A 的概率. 为了使 A 恰好出现 k 次，共有下列不同的方式：

$$\underbrace{AA\cdots A}_{k\text{个}}\underbrace{\overline{A}\cdots\overline{A}}_{n-k\text{个}} = \sigma_1$$

这表示前 k 次试验都出现 A, 后 $n-k$ 次都出现 $\overline{A}$; 或

$$\overline{A}\underbrace{A\cdots A}_{k\text{个}}\ \underbrace{\overline{A}\cdots\overline{A}}_{n-k-1\text{个}} = \sigma_2$$

这表示第一次试验出现 $\overline{A}$, 第二次直到第 $k+1$ 次都出现 A, 其余 $n-k-1$ 次都出现 $\overline{A}$; 如此继续造 $\sigma_3, \sigma_4, \cdots$, 每个 σ_i 是 E^n 的一个基本事件. 由于诸次随机试验的独立性，得

$$P(\sigma_1) = P(AA\cdots A\overline{A}\cdots\overline{A}) = p\cdot p\cdot\cdots\cdot p\cdot q\cdot\cdots\cdot q = p^k q^{n-k},$$

同理，

$$P(\sigma_i) = p^k q^{n-k}, \ \ i = 1,2,\cdots. \tag{6}$$

注意诸事件 σ_i 互不相交，而且 $(\xi = k) = \bigcup\limits_i \sigma_i$ 故

$$P(\xi = k) = P(\sigma_1) + P(\sigma_2) + \cdots. \tag{7}$$

由 (6)(7) 可见，为了求 $P(\xi = k)$, 只要求出共有多少个这种 σ_i. 注意每个 σ_i 中共含 n 个“坐位”— 字母，其中 A 占据 k 个，因此要问共有多少 σ_i, 等于问从 n 个坐位中任选 k 个，共有多少种选法 (因为这些选法与诸 σ_i 是一一对应的). 由组合论 (§1.2) 知，选法的种数是 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. 于是由 (7)(6) 得

$$\begin{aligned} P(\xi = k) &= P(\sigma_1) + P(\sigma_2) + \cdots + P\left(\sigma_{\binom{n}{k}}\right) \\ &= \binom{n}{k} p^k q^{n-k}, \ \ (k = 0,1,\cdots,n). \end{aligned} \tag{8}$$

注意 $\binom{n}{k} p^k q^{n-k}$ 是多项式 $(q+pz)^n$ 中 z^k 的系数.

例 1 (i) 设每部机床在每一分钟内需要修理的概率为 $p, 0<p<1$, 试求 n 部机床在同一分钟内有 $k(0 \leqslant k \leqslant n)$ 部需要修理的概率 p_k.

(ii) 一工人同时看管 n 部机床，由于在同一时刻只能修理一部，因此自然希望在同一分钟内有一部以上的机床需要修理的概率不能太大，譬如不超过 1%; 试问在这要求下，一工人最多能看管几部机床？

解 (i) 在一分钟内观察一部机床是否要修理是一随机试验 E. 以 A 表事件 "要修理",$P(A)=p$. 一分钟内同时观察 n 部相当于把 E 做 n 次，假定这些机床是否要修理是互相独立的，于是得一 n 重贝努里试验 $\widetilde{E}=E^n$. 以 ξ 表这 n 次试验中 A 出现的次数，亦即在同一分钟内需要修理的机床的部数，由 (8) 得

$$p_k=P(\xi=k)=\binom{n}{k}p^k q^{n-k}, \quad (k=0,1,\cdots,n).$$

(ii) 在一分钟内有一部以上机床需要修理的概率为

$$P(\xi>1)=1-P(\xi=0)-P(\xi=1)=1-q^n-npq^{n-1}.$$

令 $n_0=\max(n: \ 1-q^n-npq^{n-1} \leqslant 1\%)$, 则 n_0 是所求的机床数.

例 2 [1] **(还原抽球问题)** 设箱内共有个 N 球，其中有 N_1 个白球， $N_2=N-N_1$ 个黑球，每次自箱内等可能地取出一球，取后还原，共取 n 次，试求恰好取得 k 个白球的概率 p_k.

解 把每次抽球看成一随机试验，以 A 表 "取得白球",$\overline{A}$ 表 "取得黑球", 则 $P(A)=\frac{N_1}{N}$, $P(\overline{A})=\frac{N_2}{N}$. 还原地取 n 次球是一 n 重贝努里试验. 故

$$p_k=\binom{n}{k}\left(\frac{N_1}{N}\right)^k\left(\frac{N_2}{N}\right)^{n-k}. \tag{9}$$

例 3 (不还原抽球问题) 假设与例 2 相同，只是每次取出的球不还原，试求连取 n 次 (这相当于一次同时取 n 个球) 中取得 k 个白球的概率 r_k $(n \leqslant N)$.

解 容易看出 [2]

$$r_k=\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}, \tag{10}$$

因为取得的白球数不外是 0 或 1 或 2 $\cdots$ 或 n, 故 $\sum\limits_{k=0}^{n} r_k=1$. 因此

$$\begin{pmatrix} 0 & 1 & 2 & \cdots & n \\ r_0 & r_1 & r_2 & \cdots & r_n \end{pmatrix}$$

决定一分布，称为超几何分布；其中 r_k 由 (10) 式定义.

直观地容易想象：如果箱内球数 N 很大，那么还原抽球与不还原抽球应该相差很小，这一思想引导到下列事实：当 $N \to \infty$ 时，在一定条件下超几何分布趋于二项分布.

[1] 更一般的抽球问题见 §2.6 例 4

[2] 由 (10) 定义的超几何分布有数学期望为 $\frac{nN_1}{N}$, 方差为 $\frac{N_1(N-N_1)n(N-n)}{N^2(N-1)}$; 所以叫做超几何分布是因为 r_k 可看成为所谓超几何级数一般项的系数，正如二项分布可看成二项式 $(p+q)^n$ 的一般项一样 (见 [31] 卷一， 127 页)

精确些说，就是下面的事实：

设 $\lim\limits_{N\to\infty}\dfrac{N_1}{N}=p>0,\ \lim\limits_{N\to\infty}\dfrac{N_2}{N}=q>0,\ N_1+N_2=N,$

则对任一正整数 n, 有

$$\lim_{N\to\infty}\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}=\binom{n}{k}p^kq^{n-k},\ (k=0,1,\cdots,n). \tag{11}$$

实际上，

$$\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}=\binom{n}{k}\left(\frac{N_1}{N}\right)^k\left(\frac{N_2}{N}\right)^{n-k}$$
$$\times\frac{\left(1-\frac{1}{N_1}\right)\left(1-\frac{2}{N_1}\right)\cdots\left(1-\frac{k-1}{N_1}\right)\left(1-\frac{1}{N_2}\right)\left(1-\frac{2}{N_2}\right)\cdots\left(1-\frac{n-k-1}{N_2}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}.$$

由条件知：当 $N\to\infty$ 时， $N_1\to\infty,N_2\to\infty$; 故于上式中令 $N\to\infty$ 即得证 (11).

例 4 考虑可列重贝努里试验 $\widetilde{E}=E^\infty$, 以 η 表 A 首次出现的试验次数. 例如，对 $\widetilde{\omega}=\overline{A}\,\overline{A}\,A\overline{A}\,\overline{A}\cdots$ 则 $\eta(\widetilde{\omega})=3$, 因为 A 首次出现在第三次试验上. 易见

$$P(\eta=k)=q^{k-1}p,\ k=1,2,\cdots.$$

密度矩阵为

$$\begin{pmatrix}1 & 2 & \cdots\\ p_1 & p_2 & \cdots\end{pmatrix},\ \text{其中}p_k=q^{k-1}p \tag{12}$$

的分布叫做几何分布.

例 5 (广义贝努里试验) 设随机试验 E 只有两个基本事件 $A,\overline{A}$, 将 E 独立地重复做 n 次 (或可列次), 但在第 k 次试验中， A 出现的概率为 p_k 不出现的概率为 q_k, $0<p_k<1$, $p_k+q_k=1$. 因而 A 出现的概率与试验的次数 k 有关. 这一列试验构成广义贝努里试验，重复的次数称为试验的重数，当 $p_k=p$ 与 k 无关时，化为贝努里试验. 用证明 (8) 式同样的方法，可见在 n 重广义贝努里试验中， A 出现 k 次的概率为多项式

$$\prod_{i=1}^{n}(q_i+p_iz) \tag{13}$$

中 z^k 的系数

***例 6 (售货问题)** 某售货员同时出售两包同样的书，每次售书，他等可能地任选一包，从中取出一本，直到他某次发现一包已空为止. 问这时另一包中尚余 $r(\leqslant N)$ 本书的概率 a_r 为若干？这里 N 为每包书满装时的本数 (这问题也叫 Banach 问题).

解 每选一包构成一随机试验，共有二基本事件，概率为 $\frac{1}{2}$; 到发现一空包时，如另一包尚有 r 本，那么已做了 $2N-r$ 次试验，其中共有 N 次选自现已空的那包. 由于这 N 次可自 $2N-r$ 次中任意选出，故

$$a_r=\binom{2N-r}{N}\left(\frac{1}{2}\right)^N\left(\frac{1}{2}\right)^{N-r}=\binom{2N-r}{N}\left(\frac{1}{2}\right)^{2N-r}.\ \square$$

*(三) **二项分布 $B(n,p)$ 的一些简单性质**. 考虑 (3) 中的 $p_k(n,p)$, 当 k 从 0 变到 n 时, 试问 $p_k(n,p)$ 如何变化？

如果 $q=0$, 即 $p=1$, 由 (3) 显然得

$$p_0(n,1)=\cdots p_{n-1}(n,1)=0,\ p_n(n,1)=1.$$

以下设 $q>0$. 由 $p_k(n,p)=\binom{n}{k}p^kq^{n-k}$, 得 [1]

$$\frac{p_{k+1}(n,p)}{p_k(n,p)}=\frac{(n-k)p}{(k+1)q}.$$

当 $(n-k)p>(k+1)q$, 亦即 $np-q>k$ 时

$$p_{k+1}(n,p)>p_k(n,p); \tag{14}$$

当 $np-q=k$ 时

$$p_{k+1}(n,p)=p_k(n,p); \tag{15}$$

当 $np-q<k$ 时

$$p_{k+1}(n,p)<p_k(n,p); \tag{16}$$

分三种情况来考虑：

(i) 设 $0<np-q<n-1$. 这时 $p_k(n,p)$ 先是随 k 增大而上升，在 $k_0=[np-q]+1$ 达到极大 ($[a]$ 表不大于 a 的最大整数); 如果 $np-q=k_0-1$ 是整数，那么由 (15) 可见在 k_0-1 及 k_0 都达到极大而相等；达到极大后，由 (16) 知 $p_k(n,p)$ 就随着 k 的增大而下降. 总之，如 $np-q$ 不是整数，则

$$\begin{aligned}p_0(n.p)<\cdots<p_{k_0-1}(n,p)<p_{k_0}(n,p)\\ >p_{k_0+1}(n,p)>\cdots>p_n(n,p);\end{aligned}$$

如 $np-q$ 是整数，则

$$\begin{aligned}p_0(n.p)<\cdots<p_{k_0-1}(n,p)=p_{k_0}(n,p)\\ >p_{k_0+1}(n,p)>\cdots>p_n(n,p).\end{aligned}$$

(ii) 如果 $np-q\leqslant 0$, 由 (15)(16) 知不可能有上述上升部分；其实有

$$\begin{aligned}p_0(n,p)>p_1(n,p)>\cdots>p_n(n,p),\ \text{如}np-q<0;\\ p_0(n,p)=p_1(n,p)>\cdots>p_n(n,p),\ \text{如}np-q=0.\end{aligned}$$

---

[1] 下式给出求 $p_k(n,p)$ 的递推公式：

$$p_{k+1}(n,p)=\frac{(n-k)p}{(k+1)q}p_k(n,p);\quad p_0(n,p)=q^n$$

对许多其他离散分布，也可求出递推公式；例如，对普阿松分布见 §2.4 (7).

(iii) 如果 $np-q \geqslant n-1$, 由 (14)(15) 知不可能有下降部分，其实有

$$p_0(n,p) < \cdots < p_{n-1}(n,p) < p_n(n,p), \text{ 如 } np-q > n-1;$$
$$p_0(n,p) < \cdots < p_{n-1}(n,p) = p_n(n,p), \text{ 如 } np-q = n-1.$$

由于任一离散分布被它的密度矩阵所决定，以后为简便计，以密度矩阵来代替此分布. 考虑任一离散分布

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix},$$

如果 $\sup(p_0, p_1, \cdots)$ 在某 p_k 达到，即如

$$p_k = \sup(p_0, p_1, \cdots). \tag{17}$$

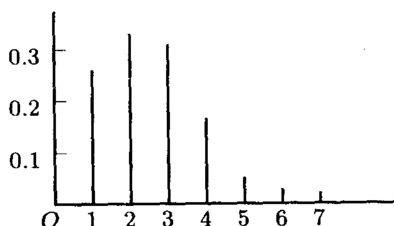
就称 a_k 为此分布的最可能值，最可能值不一定唯一.

(17) 中上确界总存在，因而最可能值至少有一个. 实际上，由于 $p_i \geqslant 0$, $\sum\limits_{i=0}^{\infty} p_i = 1$, 故至少有一 j, 使 $p_j > 0$, 对此 p_j, 存在正整数 N, 使

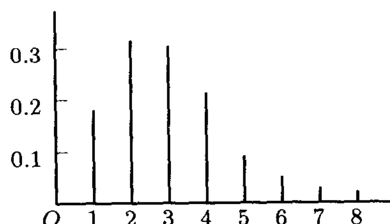
$$\sum_{i>N} p_i < p_j \tag{18}$$

于是 $$\sup(p_0, p_1, \cdots) = \max(p_0, p_1, \cdots, p_N).$$
右方有穷集的极大值总可在某 p_k 上达到.

回到二项分布情形，由上述 (i)(ii)(iii) 可见：这分布的最可能值不多于两个，当而且只当 $np-q$ 是非负整数 (包括 0) 时恰有两个，下二图说明此二情况.



$B\left(7, \frac{1}{3}\right)$ 密度矩阵图，只有一最可能值为 2.



$B\left(8, \frac{1}{3}\right)$ 密度矩阵图，有二最可能值为 2,3.

§2.4 普阿松分布与普阿松流

(一) 普阿松分布. 一种分布所以重要，通常是由于两种原因：或者它直接产生于实际问题中，或者它作为某些重要的分布的极限而出现，因而在理论上有重要意义. 我们已看到二项分布是如此，下面即将叙述的普阿松 (Poisson) 分布也是如此.

设已给常数 $\lambda > 0$, 考虑矩阵

$$\begin{pmatrix} 0 & 1 & 2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}, \quad p_k = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{1}$$

显然 $p_k > 0$, 又

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1. \tag{2}$$

故 §2.2(7) 式成立.

称密度矩阵为 (1) 的分布为普阿松分布，它含一参数 $\lambda > 0$, 记此分布为 $P(\lambda)$.

我们来证明，这分布是二项分布列的极限分布. 精确地说，考虑二项分布

$$p_k(n,p) = \binom{n}{k} p^k q^{n-k}, \quad (k = 0, 1, \cdots, n).$$

定理 1 设有一列二项分布 $\{B(n, a_n)\}$, 其中参数列 $\{a_n\}$ 满足

$$\lim_{n\to\infty} n a_n = \lambda > 0 \tag{3}$$

则对任意非负整数 k, 有

$$\lim_{n\to\infty} p_k(n, a_n) = e^{-\lambda} \frac{\lambda^k}{k!}. \tag{4}$$

证 可把 (3) 改写为 $n a_n = \lambda + o(1)$, 因而

$$a_n = \frac{\lambda}{n} + \frac{o(1)}{n}, \ 1 - a_n = 1 - \frac{\lambda}{n} - \frac{o(1)}{n} = 1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big).$$

于是

$$p_k(n, a_n) = \frac{n!}{k!(n-k)!} \Big[\frac{\lambda}{n} + \frac{o(1)}{n}\Big]^k \Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^{n-k}$$

$$\begin{aligned}
&= \frac{n!}{k!(n-k)!} \frac{[\lambda + o(1)]^k}{n^k} \frac{\Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^n}{\Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^k} \\
&= \frac{[\lambda + o(1)]^k}{k!} \Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^n \frac{n(n-1)\cdots(n-k+1)}{n^k \Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^k} \\
&= \frac{[\lambda + o(1)]^k}{k!} \Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^n \frac{1\Big(1 - \frac{1}{n}\Big)\cdots\Big(1 - \frac{k-1}{n}\Big)}{\Big[1 - \frac{\lambda}{n} - o\Big(\frac{1}{n}\Big)\Big]^k}.
\end{aligned} \tag{5}$$

注意 k 是固定的数，故当 $n \to \infty$ 时，第一因子趋于 $\frac{\lambda^k}{k!}$, 第三因子趋于 1, 第二因子

$$\Big[1 - \frac{\lambda + o(1)}{n}\Big]^n = \bigg(\Big[1 - \frac{\lambda + o(1)}{n}\Big]^{\frac{n}{\lambda + o(1)}}\bigg)^{\lambda + o(1)} \longrightarrow e^{-\lambda}$$

定理 1 的重要特殊情形是当 (3) 化为下列等式时：

$$n a_n = \lambda > 0, \text{因而} a_n = \frac{\lambda}{n} > 0 \ (\text{一切} n = 1, 2, \cdots). \tag{6}$$

利用定理 1, 既可以用二项分布来逼近普阿松分布，也可以用普阿松分布来近似具有甚大的 n 的二项分布. 如已给 $P(\lambda)$, 只需用 $B(n, \frac{\lambda}{n})$ 去逼近它；如已给 $B(n, a)$, 近似的普

阿松分布可取为 $P(na)$. 不能低估定理 1 的意义，如果 n 很大，对给定的 p, 要精确地算出 $p_k(n,p)$ 的值是很繁重的，有了定理 1, 我们可以取它的近似值 $e^{-np}\frac{(np)^k}{k!}$, 而后一值有表可查，例如

$$p_3(800,0.005)=\binom{800}{3}\cdot 0,005^3\cdot 0.995^{797}$$

的精确值为 0.1945; 由于这时 $np=800\cdot 0.005=4$, 得近似值为

$$e^{-4}\frac{4^3}{3!}=e^{-4}\cdot\frac{32}{3}=0.1954.$$

普阿松分布常被用来研究稀有事件的频数，这种事件在每次试验中出现的概率 p 很小但试验的次数 n 又很大.

例 1 [1] 自 1875 年至 1955 年中的某 63 年间，上海夏季 (即 5—9 月间) 共发生暴雨 180 次. 每年夏季共有

$$n=31+30+31+31+30=153$$

天，每次暴雨如以 1 天计算，则每天发生暴雨的概率为 $p=\frac{180}{63\times 153}$, 这值甚小，而 $n=153$ 则较大. 把暴雨看成稀有事件，对它应用普阿松分布，试求一个夏季发生 k 次 $(k=0,1,2,\cdots)$ 暴雨的概率 p_k. 先计算

$$\lambda=np=153\times\frac{180}{63\times 153}=\frac{180}{63}\approx 2.9.$$

故由 (1),$p_0=e^{-\lambda}=e^{-2.9}\approx 0.555; p_1=\lambda e^{-\lambda}=\lambda\cdot p_0=0.16$; 一般，已知 p_k 时，

$$p_{k+1}=\frac{\lambda^{k+1}}{(k+1)!}e^{-\lambda}=\frac{\lambda}{k+1}\left(\frac{\lambda^k}{k!}e^{-\lambda}\right)=\frac{\lambda}{k+1}\cdot p_k.$$

因此，63 年内，按计算应约有 $63\times p_k$ 个夏季，发生 k 次暴雨. 下面是实际观察值与理论计算值的对照表，例如，发生 2 次暴雨的，实际上有 14 个夏天，而按计算则为 14.8; 总起来看，，符合情况较好.

暴雨次数	0	1	2	3	4	5	6	7	8
实际年数	4	8	14	19	10	4	2	1	1
理论年数	3.5	10.2	14.8	14.3	10	6	2.9	1.2	0.42

图 1 表示 $B(10,\frac{3}{20})$ 与 $P(\frac{3}{2})$ 的近似误差度.

(二) p_k **的变化情况**. 为了强调 (1) 中 p_k 对 λ 的依赖性，记 p_k 为 $p_k(\lambda)$, 因而 $p_k(\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}$,

$$\frac{p_{k+1}(\lambda)}{p_k(\lambda)}=\frac{\lambda}{k+1}. \tag{7}$$

由此可见，如 $\lambda>k+1$, 则 $p_{k+1}(\lambda)>p_k(\lambda)$; 如 $\lambda=k+1$, 则 $p_{k+1}(\lambda)=p_k(\lambda)$; 如 $\lambda<k+1$, 则 $p_{k+1}(\lambda)<p_k(\lambda)$. 因此，$p_k(\lambda)$ 起初随着 k 增大而上升，在 $[\lambda]$ 达到极大，如果 $\lambda=[\lambda]$, 即 λ 是正整数时，有两个最可能

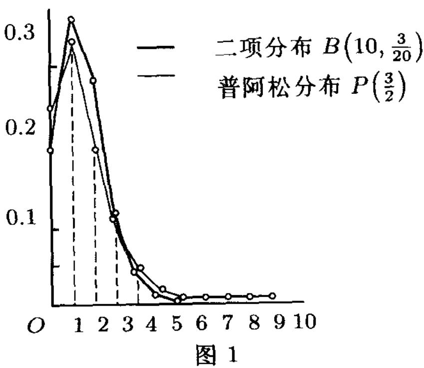


图 1

[1] 本例中数据取自么枕生著：气候统计，科学出版社，1963.

值 $\lambda-1$ 及 λ; 然后随着 k 增大而下降[1], 参看图 2.

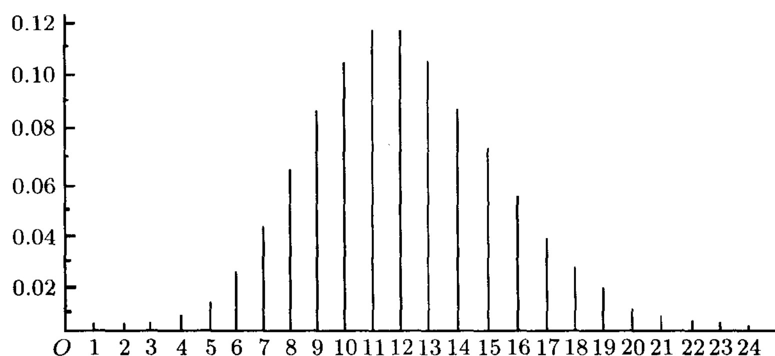


图 2 $\lambda=12$ 的普阿松分布密度矩阵图

(三) **普阿松流**. 正象贝努里试验产生二项分布一样, 普阿松流是产生普阿松分布的现实泉源之一.

源源不断地出现的许多随机的质点构成一个随机质点流, 简称为流, 例如, 在电话交换台, 要求打电话的呼唤 (每一呼唤代表一个要打电话的人) 鱼贯而来, 形成一呼唤流 (把呼唤看成质点); 到某商店去的顾客形成一顾客流; 到达某飞机场的飞机形成飞机流; 经过某块天空的流星形成流星流; 纺纱机上的断头形成断头流; 放射性物质不断放出的质点形成质点流等等. 由此可见: 随机质点流所描述的客观现象极为广泛.

以 ξ_t 表在时间区间 $(0,t]$ 内总共出现的质点个数, 我们最关心的是 ξ_t 的分布, 当流满足一些条件时, 这分布可以求出.

称流为普阿松流, 如果此流满足下列条件:

1° 在任意 n 个不相交的区间 $(a_i,b_i]$ 中, $i=1,\cdots,n$, 各自出现的质点的个数 $\xi(a_i,b_i]$ 是独立的; 也就是说, 对任意 n 个非负整数 k_i, 诸事件 $\xi(a_i,b_i]=k_i, i=1,2,\cdots,n$ 是独立的.

2° 在长为 t 的区间 $(a,a+t]$ 中, 出现 k 个质点的概率 $v_k(t)=P(\xi(a,a+t]=k)$ 与 a 无关; 而且 $v_0(t)$ 不恒等于 1.

3° 在有限区间 $(a,a+t]$ 中只出现有穷多个质点, 即

$$\sum_{k=0}^{\infty} v_k(t)=1$$

4° 在 $(a,a+t]$ 中出现一个以上质点的概率 $\psi(t)(=1-v_0(t)-v_1(t))$ 关于 t 是高级无穷小, 即

$$\lim_{t\to 0}\frac{\psi(t)}{t}=0.$$

定理 2 对于普阿松流, ξ_t 有参数为 λt 的普阿松分布, λ 是正常数; 即

$$p(\xi_t=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\ (k=0,1,2,\cdots). \tag{8}$$

证 对 $0\leqslant s<t$, 令 $v_k(t-s)=P(\xi_t-\xi_s=k)$. 由 1°, 2°,

$$v_0(1)=\left[v_0\left(\frac{1}{n}\right)\right]^n.$$

[1]如要计算 $\sum\limits_{k=0}^{m} p_k$ 的值, 可利用恒等式: $e^{-\lambda}\sum\limits_{k=0}^{m}\frac{\lambda^k}{k!}=\frac{1}{m!}\int_{\lambda}^{\infty} z^m e^{-z}\,dz$. 此式可用归纳法及分部积分证明. 而 $\int_{\lambda}^{m} z^m e^{-z}\,dz=\left(\int_0^{\infty}-\int_0^{\lambda}\right)z^m e^{-z}\,dz=m!-\int_0^{\lambda} z^m e^{-z}\,dz$; 故 $\sum\limits_{k=0}^{m} p_k=1-\frac{1}{m!}\int_0^{\lambda} z^m e^{-z}\,dz$. 最后一积分称为不完全 Gamma 函数, 其值有表可查, 参看 §8.1 中的注, 见 298 页.

设 $v_0(1)=\theta$, 则 $v_0(\frac{1}{n})=\theta^{\frac{1}{n}}$. 对任意给定的 t, 取整数 i, 使满足 $\frac{i-1}{n}<t\leqslant\frac{i}{n}$, 因而 $i=i(n)$ 与 n 有关. 令 $n\to\infty$, 由 i 的定义得 $\frac{i}{n}\to t$, $\frac{i-1}{n}\to t$. 根据 $v_0(1)$ 的概率意义, 可见 $v_0(t)$ 是 t 的不增函数, 于是

$$\theta^{\frac{i-1}{n}}=v_0\Big(\frac{i-1}{n}\Big)\geqslant v_0(t)\geqslant v_0\Big(\frac{i}{n}\Big)=\theta^{\frac{i}{n}},$$

于此式中令 $n\to\infty$, 得

$$v_0(t)=\theta^t. \tag{9}$$

注意 $0\leqslant\theta=v_0(1)\leqslant 1$, θ 与 t 无关, 故 (9) 对任意 $t\geqslant 0$ 成立.

如果说 $\theta=0$, 则 $v_0(t)=0$, 由 $1°, 2°$ 可推知在任一有限区间中出现无穷多个质点的概率为 1, 这与 $3°$ 矛盾. 如果说 $\theta=1$, 则 $v_0(t)\equiv 1$ 而与 $2°$ 矛盾, 从而 $0<\theta<1$. 故必存在常数 $\lambda>0$, 使 $\theta=e^{-\lambda}$. 由 (9) 得

$$v_0(t)=e^{-\lambda t}. \tag{10}$$

下面计算 $v_k(t)$, 它是在 $(a,a+t]$ 中出现 k 个质点的概率, $(k>0)$. 由 $2°$ 不妨取 $a=0$, 将 $(0,t]$ 分成 n 个等长的互不相交的子区间, $n>k$, 以 A_1 表事件 "在某 k 个子区间中, 各恰好出现一个质点, 而且其余 $n-k$ 个子区间内不出现质点";A_2 表 "至少有一子区间内出现一个以上的质点";A_3 表 "在 $(0,t]$ 中出现 k 个质点". 则

$$v_k(t)=P(A_1)+P(A_2A_3). \tag{11}$$

令 $\delta=\frac{t}{n}$, 有

$$P(A_1)=\binom{n}{k}[v_1(\delta)]^k[v_0(\delta)]^{n-k}.$$

由 (10) 及 $4°$, 当 $n\to\infty$ 因而 $\delta\to 0$ 时

$$\begin{aligned}
[v_0(\delta)]^{n-k}&=e^{-\lambda\delta(n-k)}=e^{-\lambda t}e^{k\lambda\delta}=e^{-\lambda t}[1+o(1)],\\
[v_1(\delta)]^k&=[1-e^{-\lambda\delta}-\psi(\delta)]^k=[1-e^{-\lambda\delta}+o(\delta)]^k\\
&=(\lambda\delta)^k[1+o(1)]=\frac{(\lambda t)^k}{n^k}[1+o(1)],
\end{aligned}$$

故

$$P(A_1)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}[1+o(1)]. \tag{12}$$

令 $n\to\infty$ 即得

$$P(A_1)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}. \tag{13}$$

其次, (11) 中 $P(A_2A_3)\leqslant P(A_2)$, 既然在任一固定的区间出现一个以上质点的概率为 $\psi(\delta)$, 由 $4°$

$$P(A_2)\leqslant n\psi(\delta)=t\frac{\psi(\delta)}{\delta}\to 0\ \ (n\to\infty).$$

由此及 (11)(12), 立得

$$P(\xi_t=k)=v_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}.\ \ \square$$

注 1. 以 η 表普阿松流中第一个质点出现的时刻，那么 $(\eta > t)$ 重合于 $(\xi_t = 0)$, 由 (10) 得

$$P(\eta > t) = P(\xi_t = 0) = v_0(t) = e^{-\lambda t}. \tag{14}$$

故如 $t > 0$, 则 $P(\eta \leqslant t) = 1 - e^{-\lambda t}$; 如 $t \leqslant 0$, 则因 η 不可能为负或 0, 故 $P(\eta \leqslant t) = 0$, 这说明 η 有指数分布 (参看 §2.2(19), 取那里的参数 $b = 0$, $a = \lambda$).

§2.5 正态分布

(一) **产生正态分布的实际背景**. 正态分布在概率论中起着非常重要的作用，在各种分布中，它居于首要的地位. 我们在实际中常常遇到一些变量，它们的分布近似于正态分布. 例如，在同一生产条件下制造的电灯泡，使用时数 ξ 随着每个灯泡而不同，譬如说，第一个可用 1200 小时，第二个 1280 小时等等，因此，ξ 是一个变量. 实践证明，ξ 的分布是近似正态的，俗话说：中间大，两头小，就是正态分布的一个性质. 一般说来，在生产条件不变的前题下，许多产品的某些量度 (如青砖的抗压强度，细纱的强力，螺丝的口径等等), 都近似地有正态分布. 这种情况在许多自然科学中也出现. 例如，热力学中理想气体分子的速度分量，射击时命中位置对目标沿某些坐标轴的偏差，物理学中测量同一物体的测量误差，生物学中同一种生物机体的某一量度 (如身长，体重), 某地区一年中的降水量等等，都是如此.

上述各种量有一共同特点：它们可以看成为许多微小的，独立的随机因素的总后果. 例如，灯泡的使用时数受着原料，工艺，保管条件等等因素的影响，而每种因素，在正常情况下，都不能起压倒一切的主导作用. 以后会看到，具有这种特点的变量一般都可认为有正态分布. 这一结论的准确的数学叙述见 §3.4 中的中心极限定理.

(二) **正态分布的定义**. 设已给二常数 a 及 $\sigma > 0$, 定义函数

$$\varphi_{a,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}, \ (x \in R_1) \tag{1}$$

它含二参数 a, σ. 特别，当 $a = 0, \sigma = 1$ 时，简记

$$\varphi(x) = \varphi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{2}$$

显然 $\varphi_{a,\sigma}(x)$ 取正值. 试证 §2.2(15) 式成立，即有

$$\int_{-\infty}^{\infty} \varphi_{a,\sigma}(x)\, dx = 1. \tag{3}$$

其实，令 $\frac{x-a}{\sigma} = y$, 则

$$\int_{-\infty}^{\infty} \varphi_{a,\sigma}(x)\, dx = \int_{-\infty}^{\infty} \varphi(y)\, dy. \tag{4}$$

然而

$$\begin{aligned}\left(\int_{-\infty}^{\infty} \varphi(y)\, dy\right)^2 &= \frac{1}{2\pi}\left(\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dxdy.\end{aligned}$$

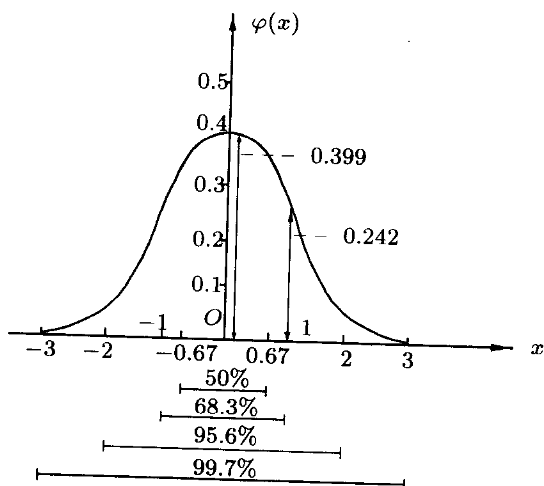


图 1 正态密度函数 $\varphi(x)$ 的图. 最下一横线表示: 在 $(-3,3)$ 中曲线下之面积为 99.7%, 其它横线意义类似. 注意全面积为 1

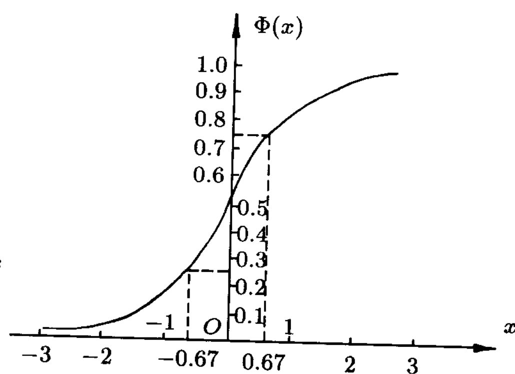


图 2 正态分布函数 $\Phi(x)$ 图

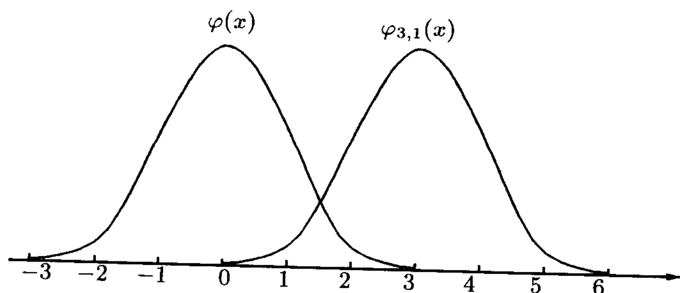


图 3

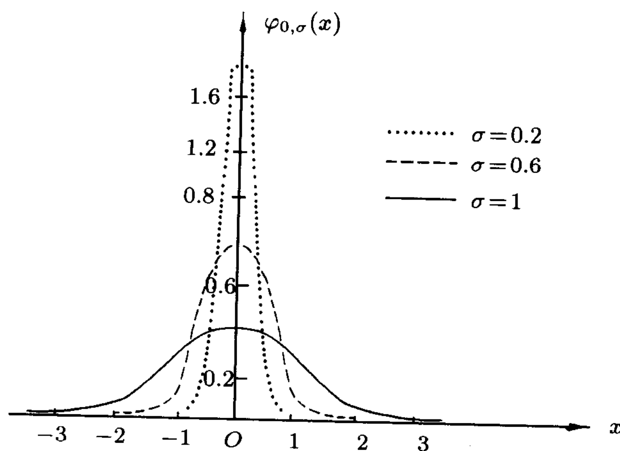


图 4

采用极坐标，令 $x = r\cos\theta, y = r\sin\theta$ 并利用

$$\int_0^\infty e^{-r^2/2}\, r dr = -e^{-r^2/2}\Big|_0^\infty = 1, \tag{5}$$

即得

$$\left(\int_{-\infty}^\infty \varphi(y)\, dy\right)^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^\infty e^{-r^2/2} r dr d\theta = 1. \tag{6}$$

既然 $\int_{-\infty}^\infty \varphi(y)\, dy$ 非负，故有上式及 (4) 即得证 (3).

称密度函数为 $\varphi_{a,\sigma}(x)$ 的分布为正态分布，为了标明二参数 $a, \sigma > 0$, 通常记此分布为 $N(a,\sigma)$, 它的分布函数记为 $\Phi_{a,\sigma}(x)$:

$$\Phi_{a,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^x e^{-(y-a)^2/2\sigma^2}\, dy,\ (x \in R_1). \tag{7}$$

$\Phi_{0,1}(x)$ 简记为 $\Phi(x)$, 即

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-y^2/2}\, dy. \tag{8}$$

(三) **简单性质**. $\varphi_{a,\sigma}(x)$ 及 $\Phi_{a,\sigma}(x)$ 具有下列性质：

1° 它们处处大于 0, 而且具有各级连续的导函数.

2° $\varphi_{a,\sigma}(x)$ 在 $(-\infty, a)$ 中严格上升，在点 a 达到极大值 $\frac{1}{\sigma\sqrt{2\pi}}$, 在 (a,∞) 严格下降；当 $x\to\infty$ 或 $x\to-\infty$ 时， $\varphi_{a,\sigma}(x)\to 0$.

3° $\varphi_{a,\sigma}(x)$ 关于点 a 对称，即

$$\varphi_{a,\sigma}(a+x) = \varphi_{a,\sigma}(a-x). \tag{9}$$

4° $\Phi_{a,\sigma}(a-x) = 1 - \Phi_{a,\sigma}(a+x).$ (10)

实际上， 1°, 2°, 3° 是明显的，由于 $\varphi_{0,\sigma}(x) = \varphi_{0,\sigma}(-x)$ 及 (3), 得

$$\begin{aligned}&\frac{1}{\sigma\sqrt{2\pi}}\left[\int_{-\infty}^{-x} e^{-y^2/2\sigma^2}\, dy + \int_{-\infty}^{x} e^{-y^2/2\sigma^2}\, dy\right]\\ &= \frac{1}{\sigma\sqrt{2\pi}}\left[\int_{-\infty}^{-x} e^{-y^2/2\sigma^2}\, dy + \int_{-x}^{\infty} e^{-y^2/2\sigma^2}\, dy\right] = 1.\end{aligned}$$

在左边作变数变换 $y = z - a$, 得

$$\frac{1}{\sigma\sqrt{2\pi}}\left[\int_{-\infty}^{a-x} e^{-(z-a)^2/2\sigma^2}\, dz + \int_{-\infty}^{x+a} e^{-(z-a)^2/2\sigma^2}\, dz\right] = 1,$$

此即 (10).

函数 $\varphi_{a,\sigma}(x)$ 的图可自 $\varphi_{0,\sigma}(x)$ 的图经坐标平移得到 (见图 3); 当 a 固定而 σ 变小时, $\varphi_{a,\sigma}(x)$ 的图形变得愈尖，因而在 $[a-h, a+h]$ 中曲线下面的面积愈来愈大，这说明分布的质量愈来愈集中在点 a 的附近，图 4 就 $a=0$ 时具体地说明了这种现象. 由此可见，参数 a,σ 有着鲜明的概率意义： σ 的大小表示分布的分散程度， σ 越大则分散程度也越大，而 a 则表示分布大部分质量所集中的地点，在 a 的周围团聚着大部分质量. 图 1 说明，按

$N(0,1)$ 分布的质量，有 99.7% 左右集中在 $[-3,3]$ 中．因此，尽管全直线每一区间 $[a,b]$ 内都有质量 $(\Phi(b)-\Phi(a)>0)$，但只有 0.3% 左右在 $[-3,3]$ 外．换句话说，如 ξ 是以 $N(0,1)$ 为分布的随机变量，则

$$P(-3\leqslant\xi\leqslant 3)=\frac{1}{\sqrt{2\pi}}\int_{-3}^{3}e^{-x^2/2}\,dx\approx 99.7\%. \tag{11}$$
$$P(\xi\overline{\in}[-3,3])\approx 0.3\%.$$

概率论中，除了已介绍过的二项分布、普阿松分布与正态分布外，还有许多重要的分布，读者可参看本节后面的常用分布表．这表的各项内容会逐渐为读者所理解．

*(四) **产生正态分布的实例**．上面两节中已经阐明：二项分布主要产生于贝努里试验，普阿松分布则联系于普阿松流，那么，正态分布的实际源泉又是什么呢？关于这一点，上面第一段中已作了一般的说明，它经常作为极限分布而出现．这里再举两个例，说明它还是某些随机变数的精确分布，并叙述推导出正态分布的思想和方法，这是很有启发性的．从历史上看，正态分布最初由高斯 (Gauss) 在研究误差理论时发现，因此我们也从这里开始．

例 1 (误差的分布) 设对某个量 (如溶液的浓度)x 进行 n 次独立的测量，得测量值为 x_i，因而误差为 $\varepsilon_i=x_i-x$，$(i=1,\cdots,n)$．这里 ε_i 及 x_i 为随机变数，而 x 为未知常数．设 ε_i 的密度为 $f(x_i-x)$,(我们用同一字母 x_i 表示随机变数及 f 的自变量，看上下文即可区别)，则 $\varepsilon_1,\cdots,\varepsilon_n$ 的联合密度为 $L\equiv\prod f(x_i-x)$，这里连乘号 $\prod$ 及以下的 $\sum$ 都对 $i=1,\cdots,n$ 进行．x 的值应使 L 极大，以最有利于 $x_1,\cdots,x_n$ 的出现，故 x 应满足 $\frac{dL}{dx}=0$ (这是最大似然法的思想，详见 §5.3)．由于 $\log x$ 是 x 的上升函数，所以 L 与 $\log L$ 在相同的点 x 上达到极大，故 x 也满足 $-\frac{d\log L}{dx}=\sum\frac{f'(x_i-x)}{f(x_i-x)}\equiv\sum g(x_i-x)=0$，(其中设 $g(y)=f'(y)/f(y)$)．由于测量值的平均数 $a=\sum x_i/n$ 最可能接近真值 x，故可设

$$G\equiv\sum g(x_i-a)\equiv\sum g(y_i)=0,\ \ (\text{其中}y_i=x_i-a).$$

但因 $\sum y_i=\sum x_i-na=0$，故这 n 个变量中只有 $n-1$ 个是独立的，令 $y_n=-(y_1+\cdots+y_{n-1})$．微分上式得

$$\frac{\partial G}{\partial y_i}+\frac{\partial G}{\partial y_n}\frac{\partial y_n}{\partial y_i}=0,$$

亦即

$$\frac{\partial g}{\partial y_i}=\frac{\partial g}{\partial y_n}\ \ (\text{一切}i),$$

这表示 $\partial g(y_i)/\partial y_i=c$(常数) 从而 $g(y)=cy+b$(b 为常数) 故 $G=c\sum y_i+nb$．但由上述 $G=0,\sum y_i=0$，故 $b=0$．于是 $f'(y)/f(y)\equiv g(y)=cy$，所以 $f(y)=k\exp\left(\frac{1}{2}cy^2\right)$，($k$ 为常数)．为了使 $f(y)$ 是密度，必须 $\int_{-\infty}^{\infty}f(y)\,dy=1$，故 c 必为负数，令 $c=\frac{-1}{\sigma^2}$，$(\sigma>0)$，代入后积分即得 $k=\frac{1}{\sigma\sqrt{2\pi}}$．由此推知误差 ε_i 应有 $N(0,\sigma)$ 正态分布．

例 2 (射击偏差的分布) 向平面上的靶进行射击，靶位在原点 O，射击后子弹落在 A 点，它的横坐标与纵坐标 X,Y 代表沿 x 轴及 y 轴的偏差，是两个随机变数 (见图 5)．试在下列假设下，求 X 与 Y 的分布．

(a) X 与 Y 的密度函数 $p(x)$ 及 $q(y)$ 都是连续的，而且 $p(0)q(0)>0$;

(b) X 与 Y 相互独立；

(c) X 与 Y 的联合密度函数在点 (x,y) 的值只依赖于此点与原点的距离 $r=\sqrt{x^2+y^2}$.

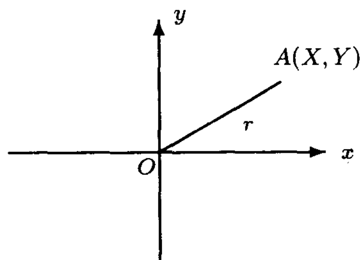


图 5

我们来证明：这时

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2};\qquad(12)$$

即 X 有 $N(0,\sigma)$ 正态分布， $\sigma>0$ 是某常数．由于 X,Y 处于平等地位，故 Y 也有 $N(0,\sigma)$ 分布．

实际上，由 (b)(c), X,Y 的联合密度在点 (x,y) 的值为

$$p(x)q(y)=s(r),\quad r=\sqrt{x^2+y^2}.\qquad(13)$$

在此式中，先令 $x=0$, 得 $p(0)q(y)=s(|y|)$; 又令 $y=0$, 得 $p(x)q(0)=s(|x|)$. 由此知 $q(y),p(x)$ 都是偶函数，于是

$$q(y)=\frac{s(|y|)}{p(0)}=\frac{p(y)q(0)}{p(0)}.$$

以此式及 $s(r)=p(r)q(0)$ 代入 (12), 并除以 $p(0)q(0)$, 即得

$$\frac{p(x)}{p(0)}\cdot\frac{p(y)}{p(0)}=\frac{p(r)}{p(0)},\quad 如 r^2=x^2+y^2.$$

令， $f(x)=\log\frac{p(x)}{p(0)}$, 改写上式为

$$f(r)=f(x)+f(y),\quad 如 r^2=x^2+y^2.$$

如果 $x^2=x_1^2+x_2^2$, 连用此式两次，得

$$f(r)=f(x)+f(y)=f(x_1)+f(x_2)+f(y),\ 如 r^2=x_1^2+x_2^2+y^2.$$

故一般地有

$$f(r)=\sum_{i=1}^{k}f(x_i),\quad 如 r^2=\sum_{i=1}^{k}x_i^2.$$

特别，选 $k=n^2$, 并令 $x=x_1=\cdots=x_k$, 有

$$f(nx)=n^2f(x);\ 或 f(n)=n^2f(1), 如 x=1.$$

又对 $x=\frac{m}{n}$, m 为整数，有

$$n^2f\left(\frac{m}{n}\right)=f\left(n\frac{m}{n}\right)=f(m)=m^2f(1),$$

即 $f\left(\frac{m}{n}\right)=c\left(\frac{m}{n}\right)^2$, 其中 $c=f(1)$; 这表示；对一切有理数 x, 有 $f(x)=cx^2$. 由假设 (a), $f(x)$ 连续，故此式对一切实数都正确．因之

$$p(x)=p(0)e^{cx^2}.$$

分布名称	分布或密度函数 $f(x)$	$f(x)$ 的图形
单点分布	$P_c=1$ (c为某常数)	
两点分布	$P_0=q,\ P_1=p,\ \left(\begin{array}{c} p\geq 0,q\geq 0 \\ p+q=1\end{array}\right)$	
二项分布 $B(n,p)$	$P_k=\binom{n}{k}p^kq^{n-k}$ $p>0,q>0$为常数,$p+q=1$ $k=0,1,\cdots,n$	
普阿松 (Poisson) 分布$P(\lambda)$	$P_k=e^{-\lambda}\frac{\lambda^k}{k!}$ $(k=0,1,\cdots,\lambda>0)$	
几何分布	$P_k=pq^{k-1},\ k=1,2,\cdots,$ $(p>0,q>0$为常数,$p+q=1)$	
均匀分布	$f(x)=\begin{cases}\frac{1}{2h}, & a-h\leq x\leq a+h;\\ 0, & 其它.\end{cases}$ (a及$h>0$为常数)	$\frac{1}{2h}$; $a-h$ a $a+h$
指数分布	$f(x)=\begin{cases}be^{-bx}, & x\geq 0;\\ 0, & x<0.\end{cases}$ $(b>0$为常数)	$b=0.5$; O 1
正态分布 $N(a,\sigma)$ $(\sigma>0)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$ (a及$\sigma>0$为常数)	$\frac{1}{\sigma\sqrt{2\pi}}$; a
$\chi^2(n)$ 分布	$f(x)=\begin{cases}0, & x\leq 0,\\ \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0.\end{cases}$ (n正整数)	0.5, 0.2, 0.1; $n=1$, $n=2$, $n=6$; O 2 4 6
加马 (Gamma) -分布$\Gamma[b,p]$	$f(x)=\begin{cases}0, & x\leq 0;\\ \frac{b^p}{\Gamma(p)}x^{p-1}e^{-bx}, & x>0.\end{cases}$ $(b>0,p>0$常数)	3/4, 1/2, 1/4; $p=1$, $p=2$, $p=4$; O 1 2 3 4

1. 设 C 为某类分布 (例如正态分布类), 如对任意 $F_1\in C$, $F_2\in C$, 它们的卷积 $F_1*F_2\in C$,(参看

分　布　表

特征函数 $\Phi(t)$	k阶矩m_k, (m_1为数学期望) k阶中心矩c_k, (c_2为方差)	附　注
e^{ict}	$m_k=c^k$ $c_k=0$	
$pe^{it}+q$	$m_k=p$ $c_2=pq$	
$(pe^{it}+q)^n$	$m_1=np$ $c_2=npq$	1. 加法定理成立[1)]： $B(n,p)*B(m,p)=B(n+m,p)$ 2. 如 ξ_i 独立，有相同的两点分布，$(i=1,\cdots,n)$则 $\sum_{i=1}^{n}\xi_i$有二项分布
$e^{\lambda(e^{it}-1)}$	$m_1=\lambda$ $c_2=\lambda$	加法定理成立 $P(\lambda_1)*P(\lambda_2)=P(\lambda_1+\lambda_2)$
$pe^{it}(1-qe^{it})^{-1}$	$m_1=p^{-1}$, $c_2=qp^{-2}$	
$e^{ita}\dfrac{\sin th}{th}$	$m_k=\dfrac{1}{2h}\dfrac{(a+h)^{k+1}-(a-h)^{k+1}}{k+1}$ $(k=1,2,\cdots)$ $c_2=\dfrac{1}{3}h^2$	如 ξ 的分布函数 $F(x)$ 连续，则 $\eta=F(\xi)$在 [0,1] 中均匀分布
$\left(1-\dfrac{it}{b}\right)^{-1}$	$m_1=\dfrac{1}{b}$ $c_2=\dfrac{1}{b^2}$	指数分布是 Gamma 分布的特殊情形
$e^{iat-\frac{\sigma^2t^2}{2}}$	各阶矩存在： $m_1=a$ $c_{2k+1}=0$ $c_{2k}=1\cdot3\cdots(2k-1)\sigma^{2k}$	加法定理成立：如 ξ_i 独立，各有分布为 $N(a_i,\sigma_i)$, d为常数 则$\sum_{i=1}^{n}c_i\xi_i+d$的分布为 $N\left(\sum_{i=1}^{n}c_ia_i+d,\sqrt{\sum_{i=1}^{n}c_i^2\sigma_i^2}\right)$
$\dfrac{1}{(1-2it)^{n/2}}$	$m_k=n(n+2)\cdots(n+2k-2)$, $(k=1,2,\cdots)$ $c_2=2n.$	1. 加法定理成立： $\chi^2(n)*\chi^2(m)=\chi^2(n+m)$. 2. 如 ξ_i 独立同分布为 $N(0,1)$, 则 $\sum_{i=1}^{n}\xi_i^2$之分布为$\chi^2(n)$
$\dfrac{1}{\left(1-\dfrac{it}{b}\right)^p}$	$m_k=\dfrac{1}{b^k}p(p+1)\cdots(p+k-1)$, $(k=1,2,\cdots)$; $c_2=\dfrac{p}{b^2}$	1. 它是 p 个 (p 为整数时) 指数分布的卷积；$p=1$ 时化为指数分布. 2. 加法定理成立： $\Gamma[b,p_1]*\Gamma[b,p_2]=\Gamma[b,p_1+p_2]$. 3. 当 $p=\dfrac{n}{2},b=\dfrac{1}{2}$ 时化为 $\chi^2(n)$ 分布

§2.8(6)), 就说对 C 加法定理成立.

分 布 名 称	分 布 或 密 度 函 数 $f(x)$	$f(x)$ 的 图 形
伯塔 (Beta) 分布	$f(x)=\begin{cases}0, & x\le 0 或 x\ge 1;\\ \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}, & 0<x<1,\end{cases}$ $(p>0,q>0$常数)	1.5 $p=q=2$ 时; o; 1
柯西 (Cauchy) 分布$c(\lambda,\mu)$	$f(x)=\dfrac{1}{\pi}\dfrac{\lambda}{\lambda^2+(x-\mu)^2}$ $(\lambda>0$常数)	$1/n\lambda$; μ
拉普拉斯 (Laplace) 分布	$f(x)=\dfrac{1}{2\lambda}e^{-\frac{\lvert x-\mu\rvert}{\lambda}}$ $(\lambda>0$常数	$\dfrac{1}{2\lambda}$; μ
学生 (Student) 分布($t(n)$分布)	$f(x)=\dfrac{1}{\sqrt{n\pi}}\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$ 其中$\Gamma(p)=\int_0^{\infty}x^{p-1}e^{-x}dx,(p>0)$	0.4; −4 −2 O 2 4; $n=3$ 时
$F-$分布 (F_{k_1,k_2})	$f(x)=\begin{cases}\dfrac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}k_1^{\frac{k_1}{2}}k_2^{\frac{k_2}{2}}\dfrac{x^{\frac{k_1}{2}-1}}{(k_2+k_1x)^{\frac{k_1+k_2}{2}}} & 如 x\ge 0;\\ 0, & 如 x<0.\end{cases}$ (k_1,k_2为正常数)	0.9; 0.6; ($k_1=10,k_2=50$); ($k_1=10,k_2=4$); 1.0; 2.0
韦布 (Weibull) 分布	$f(x)=\begin{cases}0, & (x\le 0);\\ \alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}}, & (x>0)\end{cases}$ ($\lambda>0,\alpha>0$是常数)	1; $\lambda=1$; $\alpha=3$; $\alpha=2$; O; 1; 2
对数正态分布	$f(x)=\begin{cases}0, & (x\le 0);\\ \dfrac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-(\log x-a)^2}{2\sigma^2}}, & (x>0)\end{cases}$	$\sigma=0.1$; 0.1; $\sigma=0.3$; $\sigma=0.5$; O

特征函数 $\Phi(t)$	k阶矩m_k,(m_1为数学期望) k阶中心矩c_k,(c_2为方差)	附　注
	$m_k=\dfrac{p(p+1)\cdots(p+k-1)}{(p+q)(p+q+1)\cdots(p+q+k-1)}$ $(k=1,2,\cdots)$ $c_2=\dfrac{pq}{(p+q)^2(p+q+1)}$	1. 如$\xi_1,\cdots,\xi_{n+m}$独立，同分布为$N(0,\sigma)$，则 $\sum\limits_{i=1}^{m}\xi_i^2\Big/\sum\limits_{j=1}^{n+m}\xi_j^2$ 有 Beta 分布，此时$p=\dfrac{m}{2},q=\dfrac{n}{2}$ 2. 当$p=q=\dfrac{1}{2}$时，化为反正弦分布，其密度为 $\dfrac{1}{\pi\sqrt{x(1-x)}},(0<x<1)$; 分布函数为$\dfrac{2}{\pi}\arcsin\sqrt{x}$
$e^{i\mu t-\lambda\|t\|}$	各阶矩都不存在	加法定理成立: $c(\lambda_1,\mu_1)*c(\lambda_2,\mu_2)$ $=c(\lambda_1+\lambda_2,\mu_1+\mu_2)$
$\dfrac{e^{i\mu t}}{1+\lambda^2t^2}$	各阶矩有穷	
	$k(<n)$阶矩有穷: $m_1=0,\quad (1<n)$; $m_{2v}=c_{2v}=$ $=\dfrac{1\cdot 3\cdots(2v-1)n^v}{(n-2)(n-4)\cdots(n-2v)}$, $(2v<n)$	1. 设$\xi,\xi_1,\cdots,\xi_n$独立，同分布为$N(0,\sigma)$，则 $\xi\Big/\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2}$有 Student 分布$t(n)$,(与$\sigma>0$无关). 2. $n=1$时化为 Cauchy 分布$c(1,1)$
	$m_k=\left(\dfrac{k_2}{k_1}\right)^k\dfrac{\Gamma\left(\dfrac{k_1}{2}+k\right)\Gamma\left(\dfrac{k_2}{2}-k\right)}{\Gamma\left(\dfrac{k_1}{2}\right)\Gamma\left(\dfrac{k_2}{2}\right)}$ 对$k_1<2k<k_2$存在; $c_2=\dfrac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$, $(k_2>4)$.	如 ξ,η 独立，分别有$\chi^2(k_1),\chi^2(k_2)$分布，则 $\dfrac{\xi/k_1}{\eta/k_2}$有$F-$分布$F_{k_1,k_2}$
	$m_k=\Gamma\left(\dfrac{k}{a}+1\right)\lambda^{-k/a}$, $c_2=\lambda^{-2/a}\left[\Gamma\left(\dfrac{2}{a}+1\right)-\left\{\Gamma\left(\dfrac{1}{a}+1\right)\right\}^2\right]$	当 $a=1$ 化为指数分布
	$m_1=e^{a+\sigma^2/2}$, $m_2=e^{2(a+\sigma^2)}$, $c_2=e^{2a+\sigma^2}(e^{\sigma^2}-1)$, $m_k=e^{na+(n^2\sigma^2)/2}$	设 ξ 有 $N(a,\sigma)$ 正态分布，令 $\xi=\log\eta$, 则 η 有对数正态分布

要使 $p(x)$ 在 $(-\infty,\infty)$ 上可积，必须 $c<0$, 故可令 $c=-\frac{1}{2\sigma^2}$. 再由 $\int_{-\infty}^{\infty}p(x)\,dx=1$, 得 $p(0)=\frac{1}{\sigma\sqrt{2\pi}}$. 于是得证 X 有 $N(0,1)$ 分布.

§2.6 n 维随机向量与 n 维分布

(一) 随机向量. 理论与 1 维时类似，我们只着重叙述新颖之处.

设 $\xi_1(\omega),\cdots,\xi_n(\omega)$ 为定义在同一 [1] 概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 个随机变数，它们共同构成一 n 维随机向量 $\xi(\omega)$:

$$\xi(\omega)=(\xi_1(\omega),\cdots,\xi_n(\omega)). \tag{1}$$

常把 $\xi(\omega)$ 简写为 ξ, 但留心不要与上述各节中的随机变数的记号混淆. 当 $\omega_0\in\Omega$ 固定时，$\xi(\omega_0)$ 是一通常的 n 维向量.

一维随机向量就是随机变数.

对任意 n 个实数 $x_1,\cdots,x_n$, $\omega-$ 集

$$(\xi_1(\omega)\leqslant x_1,\cdots,\xi_n(\omega)\leqslant x_n)=\bigcap_{i=1}^{n}(\xi_i(\omega)\leqslant x_i)\in\mathcal{F}; \tag{2}$$

称 n 元函数

$$F_\xi(x_1,\cdots,x_n)=P(\xi_1(\omega)\leqslant x_1,\cdots,\xi_n(\omega)\leqslant x_n) \tag{3}$$

为 $\xi(\omega)$ 的分布函数，或称为 $\xi_1(\omega),\cdots,\xi_n(\omega)$ 的联合分布函数.

引理 1 设 A 为 n 维波雷耳集，即 $A\in\mathcal{B}_n$, 则

$$(\xi(\omega)\in A)\in\mathcal{F} \tag{4}$$

证 R_n 中全体使 (4) 成立的集 A 构成 R_n 中一 $\sigma-$ 代数 $\mathfrak{M}$. 由 (2) 知 $\mathfrak{M}$ 包含一切如下 n 维区间 $\prod_{i=1}^{n}(-\infty,x_i]$, 全体这种区间的集记为 $\mathcal{G}$, 则 $\mathfrak{M}\supset\mathcal{G}$, 于是 $\mathfrak{M}\supset\sigma(\mathcal{G})=\mathcal{B}_n$(参看 §2.1), 这得证 (4). □

定义

$$P_\xi(A)=P(\xi(\omega)\in A),\quad (A\in\mathcal{B}_n). \tag{5}$$

显然 $P_\xi(A)$ 是 $\mathcal{B}_n$ 上一概率测度，称 $P_\xi(A)(A\in\mathcal{B}_n)$ 为 $\xi(\omega)$ 的分布，或称为 $\xi_1(\omega),\cdots,\xi_n(\omega)$ 的联合分布.

定理 1 函数 $F_\xi(x_1,\cdots,x_n)$ 具有下列性质：

(1°) 对任一 x_i 是单调不减的；

(2°) 对任一 x_i 是右连续的；

(3°) 令

$$F_\xi(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_n)=\lim_{x_i\to-\infty}F_\xi(x_1,\cdots,x_n),$$
$$F_\xi(\infty,\cdots,\infty)=\lim_{x_1\to\infty,\cdots,x_n\to\infty}F_\xi(x_1,\cdots,x_n),$$

则

$$F_\xi(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_n)=0,\quad F_\xi(\infty,\cdots,\infty)=1; \tag{6}$$

[1] 在以后各章节中，当同时考虑多个随机变数时，如无特别声明，总假定它们定义在同一概率空间上.

(4°) 设 $x_i \leqslant y_i$, $i=1,\cdots,n$, 则

$$F_\xi(y_1,\cdots,y_n)-\sum_{i=1}^n F_i+\sum_{i<j}F_{ij}-\cdots+(-1)^nF(x_1,\cdots,x_n)\geqslant 0, \tag{7}$$

其中 $F_{ij\cdots k}$ 是当 $z_i=x_i$, $z_j=x_j$, $\cdots$, $z_k=x_k$ 而其余 $z_l=y_l$ 时 $F_\xi(z_1,\cdots,z_n)$ 的值.

证. 性质 (1°)(2°)(3°) 可仿 §2.1 定理 1 证明，注意 (6) 的前式中是一个 $x_i\to-\infty$ 而后式中则一切 $x_i\to\infty$. 下证 (4°). 令

$$\begin{aligned}\Delta_iF_\xi(x_1,\cdots,x_n)&=F_\xi(x_1,\cdots,x_{i-1},y_i,x_{i+1},\cdots,x_n)-F_\xi(x_1,\cdots,x_n)\\&=P(\xi_j\leqslant x_j,j=1,\cdots,n,j\neq i;x_i<\xi_i(\omega)\leqslant y_i).\end{aligned}$$

由归纳法易见 [1]

$$\Delta_1\Delta_2\cdots\Delta_nF_\xi(x_1,\cdots,x_n)=P(x_i<\xi_i\leqslant y_i,i=1,\cdots,n)\geqslant 0. \tag{8}$$

然而 (8) 式左方值即 (7) 左方值，故 (7) 成立. □

注意: 性质 (4°) 不能自 (1°)(2°)(3°) 推出，当 $n=2$ 时，(7) 中左方值为 $P(\xi\in[\text{I,II,III,IV}])-P(\xi\in[\text{II,IV}])-P(\xi\in[\text{III,IV}])+P(\xi\in[\text{IV}])=P(\xi\in[\text{I}])\geqslant 0$,(见图 1).

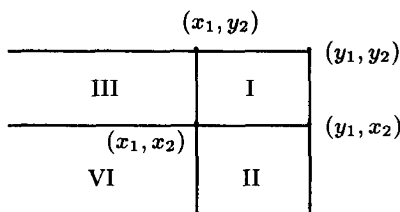


图 1

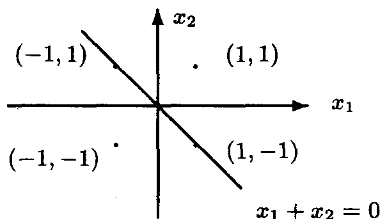


图 2

例 1 定义二元函数 (见图 2)

$$F(x_1,x_2)=\begin{cases}1, & \text{如}x_1+x_2\geqslant 0,\\ 0, & \text{如}x_1+x_2<0.\end{cases}$$

此 $F(x_1,x_2)$ 具有性质 (1°)(2°)(3°), 但

$$F(1,1)-F(-1,1)-F(1,-1)+F(-1,-1)=1-1-1+0=-1,$$

故 (4°) 不成立.

(二) **n 维分布**. 现在脱离随机向量，考虑 $(R_n,\mathcal{B}_n)$ 定义在 $\mathcal{B}_n$ 上的概率测度 $F(A)(A\in\mathcal{B}_n)$ 称为 n 维分布；定义在 R_n 上的函数 $F(x)$, $(x=(x_1,\cdots,x_n)\in R_n)$, 如果具有定理 1 中的四性质 (1°)(2°)(3°)(4°), (以 $F(x)$ 代替那里的 $F_\xi(x)$), 就称为 n 元分布函数.

***定理 2** n 维分布与 n 元分布函数是一一对应的.

证明如 §2.2 定理 1 之证，只要把那里的 $(-\infty,x]$ 改为 $\prod_{i=1}^n(-\infty,x_i]$.

[1]实际上，例如，设 $\Delta_2\cdots\Delta_nF_\xi(x_1,\cdots,x_n)=P(\xi_1\leqslant x_1,x_i<\xi_i\leqslant y_i,i=2,\cdots,n)=G(x_1)$, 则 $\Delta_1\Delta_2\cdots\Delta_nF_\xi(x_1,\cdots,x_n)=\Delta_1G(x_1)=G(y_1)-G(x_1)=P(x_i<\xi_i\leqslant y_i,i=1,\cdots,n)$

***定理 3** (存在定理) 设已给 n 元分布函数 $F(x), (x \in R_n)$, 则必存在概率空间 $(\Omega, \mathcal{F}, P)$ 及定义于其上的 n 维随机向量 $\xi(\omega)$, 使 $\xi(\omega)$ 的分布函数 $F_\xi(x)$ 满足

$$F_\xi(x) = F(x), \text{（一切} x \in R_n). \tag{9}$$

证 证明与 §2.2 定理 2 的相仿，简单重述如下：令 $(\Omega, \mathcal{F}, P) = (R_n, \mathcal{B}_n, F)$, R_n 是 n 维空间， $\omega = (\omega_1, \cdots, \omega) \in R_n$, $\mathcal{B}_n$ 是 n 维波雷耳 $\sigma-$ 代数， F 为已给的 $F(x)$ 所产生的 n 维 Lebesgue-Stieltjes 测度在 $\mathcal{B}_n$ 上的限制，满足

$$F\left(\prod_{i=1}^{n}(-\infty, x_i]\right) = F(x_1, \cdots, x_n). \tag{10}$$

定义 $\xi_i(\omega)$ 为 ω 的第 i 个坐标值，即

$$\xi_i(\omega) = \omega_i, \quad (i = 1, \cdots, n).$$

即得所欲求. 实际上，由 (10)

$$F_\xi(x_1, \cdots, x_n) = P(\xi_1 \leqslant x_1, \cdots, \xi_n \leqslant x_n) \quad = F\left(\prod_{i=1}^{n}(-\infty, x_i]\right) = F(x_1, \cdots, x_n). \quad \square$$

称分布 $F(A)(A \in \mathcal{B}_n)$ 及其对应的分布函数 $F(x_1, \cdots, x_n)$ 为离散型的，如果它集中在某个有穷或可列点集 (a_i) 上，以 $\{a\}$ 表只含一个点 a 的单点集，令 $p_i = F(\{a_i\})$, 则有

$$p_i \geqslant 0, \quad \sum_i p_i = 1$$

称 $\begin{pmatrix} a_1 & a_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$ 为密度矩阵或密度. 设点 a_i 的 n 个坐标为 $(a_{i1}, a_{i2}, \cdots, a_{in})$, 则 [1]

$$F(x_1, \cdots, x_n) = \sum p_i,$$

其中求和对一切同时满足 $a_{i1} \leqslant x_1, \cdots, a_{in} \leqslant x_n$ 的 i 进行.

称分布函数 $F(x)(x \in R_n)$ 为连续型的，如存在非负函数 $f(x)$ $(x \in R_n)$, 使

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \cdots, y_n) dy_n \cdots dy_1, \text{（一切} x \in R_n), \tag{11}$$

称 $f(x)$ 为 $F(x)$ 的密度函数或密度. 离散型的分布函数对应的分布也称为离散型的.

完全象一维时一样定义离散型与连续型的 n 维随机向量和它们的密度.

(三) **边沿分布**. 设 $F(x_1, \cdots, x_n)$ 是 n 元分布函数，对应的 n 维分布为 $F(A);(A \in \mathcal{B}_n)$. 任意保留 $k(1 \leqslant k < n)$ 个 x_i, 例如 $x_1, \cdots, x_k$, 而令其他的 x_j 都趋向 ∞, 即

$$F(x_1, \cdots, x_k, \infty, \cdots, \infty) = \lim_{x_{k+1} \to \infty, \cdots, x_n \to \infty} F(x_1, \cdots, x_n). \tag{12}$$

容易看出 $F(x_1, \cdots, x_k, \infty, \cdots, \infty)$ 是一 k 元分布函数，称为 $F(x_1, \cdots, x_n)$ 的 k 元边沿分布函数，而 $F(x_1, \cdots, x_k, \infty, \cdots, \infty)$ 所对应的 k 维分布则称为 $F(A)$ 的 k 维边沿分布. 由于自 $x_1, \cdots, x_n$ 中挑选 k 个 x_i 的方法共有 $\binom{n}{k}$ 种，故共有 $\binom{n}{k}$ 个 k 维边沿分布函数.

---

[1] 一般地，对任意 $A \in \mathcal{B}_n$, 有 $F(A) = \sum_{(i:a_i \in A)} p_i$; 对连续型则为 $F(A) = \int \cdots \int_A f(y_1, \cdots, y_n) dy_1 \cdots dy_n$

如果 $F(x_1,\cdots,x_n)$ 是连续型的，有密度为 $f(x_1,\cdots,x_n)$, 那么

$$\begin{aligned}&F(x_1,\cdots,x_k,\infty,\cdots,\infty)\\=&\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(y_1,\cdots,y_n)dy_n\cdots dy_k\cdots dy_1.\end{aligned}\tag{13}$$

可见 $F(x_1,\cdots,x_k,\infty,\cdots,\infty)$ 也是连续型的，密度为

$$g(x_1,\cdots,x_k)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)dx_n\cdots dx_{k+1}.\tag{14}$$

如 $F(x_1,\cdots,x_n)$ 是离散型的，那么 $F(x_1,\cdots,x_k,\infty,\cdots,\infty)$ 也是离散型的，其密度可仿 (14) 求得，只须把积分号改为求和号. 作为简单的例，设二维密度

$$\begin{pmatrix}(1,1) & (2,1) & (2,2) & (3,4)\\ \frac{1}{3} & \frac{3}{24} & \frac{1}{24} & \frac{1}{2}\end{pmatrix}$$

决定分布函数 $F(x_1,x_2)$, 其中四个点的横坐标构成集 (1,2,3). 仿 (14), 可见 $F(x_1,\infty)$ 的密度为

$$\begin{pmatrix}1 & 2 & 3\\ \frac{1}{3} & \frac{3}{24}+\frac{1}{24} & \frac{1}{2}\end{pmatrix}=\begin{pmatrix}1 & 2 & 3\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2}\end{pmatrix}$$

类似地， $F(\infty,x_2)$ 的密度为

$$\begin{pmatrix}1 & 2 & 4\\ \frac{11}{24} & \frac{1}{24} & \frac{1}{2}\end{pmatrix}$$

如果 n 维随机向量 $(\xi_1,\cdots,\xi_n)$ 的分布函数是 $F(x_1,\cdots,x_n)$, 即如

$$F(x_1,\cdots,x_n)=P(\xi_1\leqslant x_1,\cdots,\xi_n\leqslant x_n),$$

令 $x_i\to\infty$, $i=k+1,\cdots,n$, 则

$$F(x_1,\cdots,x_k,\infty,\cdots,\infty)=P(\xi_1\leqslant x_1,\cdots,\xi_k\leqslant x_k).$$

因而 $(\xi_1,\cdots,\xi_k)$ 的分布函数为 $F(x_1,\cdots,x_k,\infty,\cdots,\infty)$. 由此可见，如 $(\xi_1,\cdots,\xi_n)$ 是连续型 (离散型) 的，那么 $(\xi_1,\cdots,\xi_n)$ 也是连续型 (离散型) 的；后者的密度在上面已求出.

以上我们对保留 $x_1,\cdots,x_k$ 的情况进行了讨论，如保留的是 $x_{i_1},\cdots,x_{i_k}$, 讨论完全类似.

边沿分布函数由分布函数 $F(x_1,\cdots,x_n)$ 唯一决定；但反之不然，也就是说，不同的分布函数可以有一切相同的边沿分布函数.

例 2 设有两个二元分布函数为 $F(x,y)$ 及 $G(x,y)$, 分别有密度函数为

$$f(x,y)=\begin{cases}x+y, & \text{如}0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1;\\ 0, & \text{反之}.\end{cases}$$

$$g(x,y)=\begin{cases}\left(\dfrac{1}{2}+x\right)\left(\dfrac{1}{2}+y\right), & \text{如}0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1;\\ 0, & \text{反之}.\end{cases}$$

容易看出，方程

$$\left(\frac{1}{2}+x\right)\left(\frac{1}{2}+y\right)=x+y$$

只有根为 $x=\frac{1}{2}, y=\frac{1}{2}$, 因而在正方形 $0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1$ 中，只是沿二直线 $x=\frac{1}{2}$ 及 $y=\frac{1}{2}$ 上，$f(x,y)=g(x,y)$, 可见 $F(x,y)$ 与 $G(x,y)$ 不恒等. 然而两对边沿分布函数恒等，因为它们的两对密度函数相等的缘故. 实际上：

$$\begin{aligned}\int_{-\infty}^{\infty} f(x,y)\,dy &= \int_0^1 (x+y)\,dy = \frac{1}{2}+x\\ &= \int_0^1 \left(\frac{1}{2}+x\right)\left(\frac{1}{2}+y\right)dy = \int_{-\infty}^{\infty} g(x,y)\,dy.\\ \int_{-\infty}^{\infty} f(x,y)\,dx &= \int_0^1 (x+y)\,dx = \frac{1}{2}+y\\ &= \int_0^1 \left(\frac{1}{2}+x\right)\left(\frac{1}{2}+y\right)dx = \int_{-\infty}^{\infty} g(x,y)\,dx.\end{aligned}$$

例 3 设随机试验 E 有 m 个基本事件 $A_1, A_2, \cdots, A_m$, $P(A_i)=p_i>0$ 因而 $\sum\limits_{i=1}^{m} p_i=1$. 今将 E 独立地重复 n 次而得 $\widetilde{E}=E^n$, 以 ξ_i 表此 n 次试验中 A_i 出现的次数，仿照 §2.3(8) 的证，可见

$$P(\xi_1=k_1, \xi_2=k_2, \cdots, \xi_m=k_m)=\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}, \tag{15}$$

$k_i\geqslant 0$ 为整数，$\sum\limits_{i=1}^{m} k_i=n$. $m=2$ 时，$\widetilde{E}$ 为 n 重贝努里试验.

例 4 (一般的抽球问题) 设箱内共有 $N=\sum\limits_{j=1}^{m} N_j$ 个球，其中 N_j 表第 j 种颜色的球的个数. 今自箱内接连地取出 n 个球，试求所取出的球中，恰有第 1 种颜色的球共 k_1 个，$\cdots$ 第 m 种颜色的球共 k_m 个的概率. $\sum\limits_{j=1}^{m} k_j=n$

解 在还原情况下，所求概率为

$$\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}$$

其中 $p_j=\frac{N_j}{N}$, 此式是 §2.3(9) 中右方值的一般化.

在不还原情况下，所求的概率为

$$\frac{\binom{N_1}{k_1}\binom{N_2}{k_2}\cdots\binom{N_m}{k_m}}{\binom{N}{n}}.$$

这是 §2.3(10) 中右方值的一般化.

现在脱离随机试验，设 $k=(k_1,\cdots,k_m)$ 是一 m 维点，其中 k_i 为非负整数，而且 $\sum\limits_{i=1}^{m} k_i=n$, n 固定. 全体这样的点 k 构成 R_m 中一有穷集 Λ; 又设已给 m 个非负数 p_i, $\sum\limits_{i=1}^{m} p_i=1$. 由

$$F(\{k\})=\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}, \quad (k\in\Lambda) \tag{16}$$

定义的离散分布 F 称为 m 项分布，注意

$$\sum_{k\in\Lambda} F(\{k\}) = (p_1+p_2+\cdots+p_m)^n = 1.$$

当 $m=2$ 时它化为二项分布.

例 5 设 $D\in\mathcal{B}_n$, D 的勒贝格测度 $|D|>0$. 定义函数

$$f(x_1,\cdots,x_n)=\begin{cases}1/|D|, & \text{如}(x_1,\cdots,x_n)\in D,\\ 0, & \text{如}(x_1,\cdots,x_n)\overline{\in} D,\end{cases} \tag{17}$$

以 $f(x_1,\cdots,x_n)$ 为密度函数的分布称为在 D 上的 n 维均匀分布，或简称均匀分布，它是 §2.2 例 2 的一般化.

(四) n 维正态分布. 这是最重要的多维分布.

例 6 设已给常数 $a,b,\sigma_1>0,\sigma_2>0$ 及 $r,|r|<1$. 定义函数

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\times\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}, \tag{18}$$

它的切断了的图形见图 3. 显然 $f(x,y)>0$, 下面证明

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy=1, \tag{19}$$

以 $f(x,y)$ 为密度函数的分布 F 称为二维正态分布. 为证 (19), 令 $\frac{x-a}{\sigma_1}=u$, $\frac{y-b}{\sigma_2}=v$. 先计算

$$\begin{aligned}f_1(x)&=\int_{-\infty}^{\infty} f(x,y)\,dy=\frac{1}{2\pi\sigma_1\sqrt{1-r^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-r^2)}[u^2-2ruv+v^2]\right\}dv\\&=\frac{1}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\exp\left\{-\frac{1}{2(1-r^2)}[(v-ru)^2+(1-r^2)u^2]\right\}dv\\&=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-u^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\exp\left\{-\frac{(v-ru)^2}{2(1-r^2)}\right\}dv\end{aligned}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-u^2/2} \text{(利用§2.5(3))}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma_1^2}}; \tag{20}$$

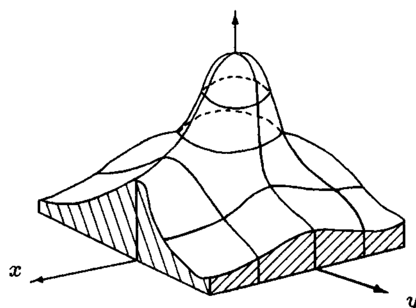


图 3

于是

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = \int_{-\infty}^{\infty} \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma_1^2}}\,dx = 1.$$

由对称性,

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(y-b)^2}{2\sigma_2^2}}. \tag{21}$$

然而 $f_1(x), f_2(y)$ 是两个边沿分布的密度函数，故二边沿分布是一维正态的，即 $N(a,\sigma_1)$ 及 $N(b,\sigma_2)$.(但反之不真：两边沿分布都是正态的，联合分布却未必是二维正态的.)(参看 144 页，题 62.)

现在考虑下列问题：取椭圆

$$\frac{(x-a)^2}{\sigma_1^2} - \frac{2r(x-a)(y-b)}{2\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} = c^2 \tag{22}$$

由 (18) 可见，当 (x,y) 在此椭圆上变动时，$f(x,y)$ 的值不变而恒等于常数

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{c^2}{2(1-r^2)}},$$

因而称椭圆 (22) 为等概率椭圆. 以 A_c 表此椭圆围成的区域，试求 $F(A_c)$, F 是以 (18) 中 $f(x,y)$ 为密度函数的二维正态分布. 直观地说：设质点 A 按二维正态律分布，则它落在椭圆 (22) 内的概率为 $F(A_c)$.

引进极坐标 $x-a=\rho\cos\theta, y-b=\rho\sin\theta$. 并令

$$s^2 = \frac{1}{1-r^2}\left[\frac{\cos^2\theta}{\sigma_1^2} - 2r\frac{\cos\theta\sin\theta}{\sigma_1\sigma_2} + \frac{\sin^2\theta}{\sigma_2^2}\right],$$

则

$$F(A_c) = \int_{A_c}\int f(x,y)\,dxdy$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^{2\pi}\int_0^{c/s\sqrt{1-r^2}} e^{-\frac{\rho^2}{2}s^2}\rho d\rho d\theta$$

$$= \frac{1-e^{-\frac{c^2}{2(1-r^2)}}}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^{2\pi} \frac{d\theta}{s^2}. \tag{23}$$

于上式中令 $c\to\infty$, 得

$$1 = \lim_{c\to\infty} F(A_c) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^{2\pi} \frac{d\theta}{s^2}.$$

因而附带求出了积分 $\int_0^{2\pi}\frac{d\theta}{s^2}$ 的值为 $2\pi\sigma_1\sigma_2\sqrt{1-r^2}$; 代入 (23) 即得

$$F(A_c)=1-e^{-\frac{c^2}{2(1-r^2)}}. \tag{24}$$

例 7 设已给 n 级矩阵 $B=(b_{jk})$, 满足条件:

(A) 对称性: $b_{jk}=b_{kj}$ (一切 $j,k=1,\cdots,n$);

(B) 正定性: 对任意实数 $\eta_1,\cdots,\eta_n$, 如果它们不全为 0, 则

$$\sum_{j,k=1}^{n} b_{jk}\eta_j\eta_k>0. \tag{25}$$

称这样的矩阵 B 为 n 级正定对称矩阵. 以 $B^{-1}=(r_{jk})$ 表 B 的逆矩阵, 以 $|B|$ 表 B 的行列式的值.

又设已给向量 $a=(a_1,\cdots,a_n)$, a_i 为任意实数, 定义 n 元函数:

$$f(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{n/2}|B|^{1/2}}\exp\Big\{-\frac{1}{2}\sum_{j,k=1}^{n}r_{jk}(x_j-a_j)(x_k-a_k)\Big\}. \tag{26}$$

用矩阵记号, 令 c' 表矩阵 c 的转置矩阵, 则 (26) 可改写为

$$f(x)=\frac{1}{(2\pi)^{n/2}|B|^{1/2}}\exp\Big\{-\frac{1}{2}(x-a)B^{-1}(x-a)'\Big\},\ (x\in R_n) \tag{27}$$

显然 $f(x)>0$, 下证

$$\int_{R_n}f(x)\,dx=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)\,dx_1\cdots dx_n=1. \tag{28}$$

称以 (27) 中的 $f(x)$ 为密度函数的连续型分布为 n 维正态分布, 特别, 当 $B=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$ 时, 化为例 6 中 2 维正态分布.

由于 B 的对称性, 根据矩阵论知存在正交矩阵 T, 使

$$TBT'=D=\begin{pmatrix}d_1 & & & 0\\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n\end{pmatrix}, \tag{29}$$

其中 D 的非对角线元素皆为 0 而 d_i 是 B 的特征值, 即方程 $|B-dI|=0$ 的根, $I=\begin{pmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{pmatrix}$, 可能 $d_i=d_j$, 由 B 的正定性知 $d_i>0$, $i=1,\cdots,n$. 作变换

$$y=(x-a)T',\ 即x=a+yT,$$

注意 $T^{-1}=T'$ 也是正交的, 由此推知变换行列式 $\frac{\partial(x)}{\partial(y)}$ 之值为 $|T|=\pm1$, (27) 中的

$$|B|=|T'DT|=|T'||D||T|=|T'T||D|=d_1\cdots d_n,$$

$$(x-a)B^{-1}(x-a)'=yTB^{-1}T'y'=y(T')^{-1}B^{-1}T^{-1}y'$$

$$=y(TBT')^{-1}y'=yD^{-1}y'=\sum_{i=1}^{n}y_i^2/d_i,$$

因此 (28) 左方为

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{n/2}(d_1\cdots d_n)^{1/2}}e^{-\sum\limits_{i=1}^{n}\frac{y_i^2}{2d_i}}dy_1\cdots dy_n$$
$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi d_i}}\int_{-\infty}^{\infty}e^{-y_i^2/2d_i}dy_i=1.$$

由于此分布依赖于 n 维向量 $a=(a_1,\cdots,a_n)$ 及 n 维正定对称矩阵 B, 故宜记此分布为 $N(a,B)$. 关于它的进一步讨论见 §2.12(三).

§2.7 随机变数的独立性；条件分布

(一) 独立性. 设 $\xi_1(\omega),\cdots,\xi_n(\omega)$ 为定义在概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 个随机变数，称它们为相互独立的，或独立的，如果对任意实数 $x_1,\cdots,x_n$, 有

$$F(x_1,\cdots,x_n)=F_1(x_1)\cdots F_n(x_n),\tag{1}$$

这里 F 是 $\xi_1,\cdots,\xi_n$ 的联合分布函数， F_i 是 ξ_i 的一元分布函数. 等式 (1) 可转述为：联合分布函数 $F(x_1,\cdots,x_n)$ 等于 n 个一维边沿分布函数的积，有时把 (1) 写成下式更方便：

$$P(\xi_1\leqslant x_1,\cdots,\xi_n\leqslant x_n)=\prod_{i=1}^{n}P(\xi_i\leqslant x_i).\tag{1'}$$

称定义在 $(\Omega,\mathcal{F},P)$ 上的随机变数列 $\{\xi_n(\omega)\}$ 是独立的，如果其中任意有限多个随机变数是独立的.

先分别考虑两类随机变数.

引理 1 设 ξ_i 是离散型随机变数, 它的分布集中在集 Λ_i 上，$(i=1,2,\cdots,n)$; 则 $\xi_1,\cdots,\xi_n$ 独立的充要条件是

$$P(\xi_1=a_1,\cdots,\xi_n=a_n)=\prod_{i=1}^{n}P(\xi_i=a_i),\tag{2}$$

对任意 $a_i\in\Lambda_i(i=1,\cdots,n)$ 成立.

证 将 (2) 式两方对一切满足 $a_1\leqslant x_1,\cdots,a_n\leqslant x_n$ 的 $(a_1,\cdots,a_n)$ 相加即得 (1'). 反之，设 (1) 成立，于是有

$$P(\xi_1\leqslant a_1,\cdots,\xi_{n-1}\leqslant a_{n-1},\xi_n\leqslant a_n)=\prod_{i=1}^{n}P(\xi_i\leqslant a_i),\tag{3}$$

$$P(\xi_1\leqslant a_1,\cdots,\xi_{n-1}\leqslant a_{n-1},\xi_n\leqslant x_n)=\prod_{i=1}^{n-1}P(\xi_i\leqslant a_i)P(\xi_n\leqslant x_n).$$

令 $x_n\uparrow a_n$, 由后式得

$$P(\xi_1\leqslant a_1,\cdots,\xi_{n-1}\leqslant a_{n-1},\xi_n< a_n)=\prod_{i=1}^{n-1}P(\xi_i\leqslant a_i)P(\xi_n< a_n).$$

自 (3) 式减去此式，即得

$$P(\xi_1 \leqslant a_1, \cdots, \xi_{n-1} \leqslant a_{n-1}, \xi_n = a_n) = \prod_{i=1}^{n-1} P(\xi_i \leqslant a_i) P(\xi_n = a_n) \tag{4}$$

对 (4) 中 a_{n-1} 作同样讨论，可得

$$P(\xi_1 \leqslant a_1, \cdots, \xi_{n-2} \leqslant a_{n-2}, \xi_{n-1} = a_{n-1}, \xi_n = a_n) = \prod_{i=1}^{n-2} P(\xi_i \leqslant a_i) P(\xi_{n-1} = a_{n-1}) P(\xi_n = a_n)$$

如此共重复 $n-1$ 次后即得 (2). □

引理 2 设 $\xi_1, \cdots, \xi_n$ 是连续型的随机变数，则它们独立的充要条件是：关于 R_n 中勒贝格测度几乎处处有

$$f(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_i(x_i), \tag{5}$$

其中 f 及 f_i 分别是 F 及 F_i 的密度函数；如 f, f_i 都是连续函数，则充要条件是 (5) 式处处成立.

证 充分性：由 (5) 得

$$\begin{aligned} F(x_1, \cdots, x_n) &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \cdots, y_n)\, dy_n \cdots dy_1 \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(y_1), \cdots, f_n(y_n)\, dy_n \cdots dy_1 = \prod_{i=1}^{n} F_i(x_i), \end{aligned}$$

必要性：由 (1) 得

$$\begin{aligned} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \cdots, y_n)\, dy_n \cdots dy_1 &= F(x_1, \cdots, x_n) \\ = \prod_{i=1}^{n} F_i(x_i) &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(y_1), \cdots, f_n(y_n)\, dy_n \cdots dy_1. \end{aligned}$$

这两边对 $x_1, \cdots, x_n$ 各微分一次，即知 (5) 式几乎处处成立.

二连续函数如几乎处处相等，则必恒等，此得证第二结论.

定理 1 $\xi_1, \cdots, \xi_n$ 独立的充要条件是：对任意 n 个一维波雷耳集 $A_1, \cdots, A_n$, 有

$$P(\xi_1 \in A_1, \cdots, \xi_n \in A_n) = \prod_{i=1}^{n} P(\xi_i \in A_i). \tag{6}$$

***证** 充分性：取 $A_i = (-\infty, x_i]$, 由 (6) 得 (1').

必要性：定义 n 元分布函数 $\widetilde{F}(x_1, \cdots, x_n) = F_1(x_1) \cdots, F_n(x_n)$, 它产生的 n 维勒贝格－司梯阶测度记为 $\widetilde{F}(A)$, $A \in \mathcal{B}_n$, 满足条件：对任意 $A_i \in \mathcal{B}_1$, $i = 1, \cdots$, 有 [1]

$$\widetilde{F}(A_1 \times \cdots \times A_n) = F_1(A_1) \cdots F_n(A_n),$$

[1] 参看 [10] 第三章 §5.

其中 $A_1\times\cdots\times A_n$ 表 Cartesian 乘积，即 R_n 中一切满足 $y_1\in A_1,\cdots,y_n\in A_n$ 的点 $(y_1,\cdots,y_n)$ 的集，而 $F_i(A)$ 是 $F_i(x)$ 产生的测度，它是 ξ_i 的分布. 由 (1) 知 $F(x_1,\cdots,x_n)=\widetilde{F}(x_1,\cdots,x_n)$，故它们产生同一测度 $\widetilde{F}(A)$, $A\in\mathcal{B}_n$. 既然 $F(x_1,\cdots,x_n)$ 是 $\xi_1,\cdots,\xi_n$ 的分布函数，故 $\widetilde{F}(A)$ 是它们的联合分布，于是

$$
\begin{aligned}
P(\xi_1\in A_1,\cdots,\xi_n\in A_n)&=\widetilde{F}(A_1\times\cdots\times A_n)=F_1(A_1)\cdots F_n(A_n)\\
&=\prod_{i=1}^{n}P(\xi_i\in A_i).\quad\square
\end{aligned}
$$

系 1 设随机变数 $\xi_1,\cdots,\xi_n$ 独立，又 $f_i(x)$ 为一元波雷耳可测函数，$i=1,\cdots,n$, 则 $f_1(\xi_1),\cdots,f_n(\xi_n)$ 也是独立的随机变数.

***证** 因 $f_i(x)$ 波雷耳可测，故对任意 $a\in R_1$, 集 $A_i^a=(x:f_i(x)\leqslant a)\in\mathcal{B}_1$; $(f_i(\xi_i)\leqslant a)=(\xi_i\in A_i^a)\in\mathcal{F}$ (§2.1 引理 3), 故 $f_i(\xi_i)$ 是随机变数. 其次

$$P(\omega:\ f_i(\xi_i(\omega))\leqslant x_i,\ i=1,\cdots,n)=P(\omega:\ \xi_i(\omega)\in A_i^{x_i},\ i=1,\cdots,n),$$

由定理 1, 后式等于

$$\prod_{i=1}^{n}P(\omega:\ \xi_i(\omega)\in A_i^{x_i})=\prod_{i=1}^{n}P(\omega:\ f_i(\xi_i(\omega))\leqslant x_i).\quad\square$$

***定理 2 (存在定理)** 设已给 n 个一维分布函数 $F_i(x)$, 则必存在概率空间 $(\Omega,\mathcal{F},P)$ 及定义于其上的独立随机变数 $\xi_1,\cdots,\xi_n$, 使 ξ_i 的分布函数重合于 $F_i(x)$, $i=1,\cdots,n$.

证 定义 n 元函数 $F(x_1,\cdots,x_n)$,

$$F(x_1,\cdots,x_n)=\prod_{i=1}^{n}F_i(x_i).\tag{7}$$

易见它是一 n 元分布函数，根据 §2.6 定理 3, 知存在概率空间 $(\Omega,\mathcal{F},P)=(R_n,\mathcal{B}_n,F)$, $\omega=(\omega_1,\cdots,\omega_n)\in R_n$, 及坐标函数 $\xi_i(\omega)=\omega_i$, $\xi_i(\omega)$ 是随机变数而且 $\xi_1(\omega),\cdots,\xi_n(\omega)$ 的联合分布函数为 $F(x_1,\cdots,x_n)$. 由此及 (7) 得

$$P(\xi_1\leqslant x_1,\cdots,\xi_n\leqslant x_n)=F(x_1,\cdots,x_n)=\prod_{i=1}^{n}F_i(x_i).\tag{8}$$

于 (8) 中令 $x_i\to\infty$, 一切 $i\neq j$, 即得

$$P(\xi_j\leqslant x_j)=F_j(x_j).$$

故知 $F_j(x)$ 是 ξ_j 的分布函数，根据 (7) 式即知 $\xi_1,\cdots,\xi_n$ 是独立的. $\square$

定理 2 的结论对已给的一列分布函数 $F_i(x), i=1,2,\cdots$ 也正确，不过证明中涉及无穷维空间中的测度问题，故从略.

例 1 设 (ξ_1,ξ_2) 有二维正态分布，密度函数 $f(x,y)$ 由 §2.6(18) 定义，即

$$
\begin{aligned}
f(x,y)=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\Big\{-\frac{1}{2(1-r^2)}\\
&\times\Big[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\Big]\Big\}.
\end{aligned}\tag{9}
$$

在 §2.6 中已证明：ξ_1,ξ_2 各有一维正态分布 $N(a,\sigma_1)$, $N(b,\sigma_2)$, 密度分别为

$$f_1(x)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma_1^2}},f_2(y)=\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(y-b)^2}{2\sigma_2^2}}. \tag{10}$$

由引理 2 立得：为使 ξ_1,ξ_2 独立，充分条件是 $r=0$.

以后 (§2.10) 会看到，这条件也是必要的.

例 2 在贝努里试验中，定义 $\xi_i=1$, 如 A 在第 i 次试验中出现. 否则令 $\xi_i=0$. 由试验的独立性知 $\{\xi_i\}$ 是一列独立的随机变数，这时概率空间应取为乘积空间 $\widetilde{\Omega}=\Omega^\infty$(参看 §2.3),$\Omega=(A,\overline{A})$. ξ_i 只依赖于 $\widetilde{\omega}$ 的第 i 个因子. 例如

$$\xi_i(\underbrace{A\cdots A}_{i-1\text{ 个}}A\overline{A}\cdots)=1, \tag{11}$$

$$\xi_i(\underbrace{A\cdots A}_{i-1\text{ 个}}\overline{A}\,\overline{A}\cdots)=0, \tag{12}$$

$$P(\xi_1=k_1,\cdots,\xi_n=k_n)=\prod_{i=1}^{n}P(\xi_i=k_i)=p^mq^{n-m},$$

其中 $k_i=0$ 或 $1,p$ 为在一次试验中 A 出现的概率，$q=1-p,m$ 是等于 1 的 k_i 的个数.

(二) **条件分布**. 先回忆条件概率的定义

$$P(A|B)=\frac{P(AB)}{P(B)},\quad 如P(B)>0. \tag{13}$$

设已给二随机变数 ξ,η, 对任意 $C\in\mathcal{B}_1$, 如 $P(\xi\in C)>0$, 可考虑 $y\in R_1$ 的函数

$$P(\eta\leqslant y|\ \xi\in C)=\frac{P(\eta\leqslant y,\xi\in C)}{P(\xi\in C)}. \tag{14}$$

容易看出 $P(\eta\leqslant y|\ \xi\in C)$ 是一维分布函数，自然的称它为在条件 $\xi\in C$ 下，η 的条件分布函数，它所对应的分布称为在条件 $\xi\in C$ 下，η 的条件分布.

在许多问题中常常要考虑 C 是单点集 $\{x\}$ 的情形，如果 $P(\xi=x)>0$, 那么这没有任何困难，问题发生在当 $P(\xi=x)=0$ 时，应该如何定义 $P(\eta\leqslant y|\xi=x)$ 才合理. 注意，当 ξ 有连续型分布时，一定有 $P(\xi=x)=0$; 实际上，设 $f_\xi(y)$ 是 ξ 的分布密度函数，那么

$$P(\xi=x)=\lim_{n\to\infty}P\Big(x\leqslant\xi<x+\frac{1}{n}\Big)=\lim_{n\to\infty}\int_x^{x+\frac{1}{n}}f_\xi(y)\,dy=0. \tag{15}$$

逐步考虑各种类型的随机变数.

(甲) 设 ξ,η 是离散型的.

$$\xi=\begin{pmatrix}x_0 & x_1 & x_2 & \cdots\\ p_0 & p_1 & p_2 & \cdots\end{pmatrix},\quad \eta=\begin{pmatrix}y_0 & y_1 & y_2 & \cdots\\ q_0 & q_1 & q_2 & \cdots\end{pmatrix}.$$

不妨设一切 $p_i>0,q_i>0$; 令

$$P(\xi=x_i,\eta=y_j)=p_{ij}. \tag{16}$$

于是

$$p_i=\sum_j p_{ij},\quad q_j=\sum_i p_{ij}, \tag{17}$$

从而

$$P(\eta=y_j|\xi=x_i)=\frac{p_{ij}}{p_i}=\frac{p_{ij}}{\sum\limits_k p_{ik}}, \tag{18}$$

$$P(\eta\leqslant y|\xi=x_i)=\frac{\sum\limits_{j:y_j\leqslant y} p_{ij}}{\sum\limits_k p_{ik}}. \tag{19}$$

(乙) 设 ξ,η 是连续型的. 以 $f(x,y)$ 表 (ξ,η) 的分布密度，如果在定点 x,

$$\int_{-\infty}^{\infty} f(x,y)\,dy>0 \tag{20}$$

由于 (19)(18) 的启发，自然地定义

$$P(\eta\leqslant y|\xi=x)=\frac{\int_{-\infty}^{y} f(x,z)\,dz}{\int_{-\infty}^{\infty} f(x,z)\,dz} \tag{21}$$

这定义的好处是避免了 $P(\xi=x)=0$ 的困难. 称 y 的函数 $f(y|x)$

$$f(y|x)=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,z)\,dz} \tag{22}$$

为在条件 $\xi=x$ 下 η 的条件分布密度. 注意 $\int_{-\infty}^{\infty} f(x,z)\,dz$ 是 (ξ,η) 的边沿密度，也就是 ξ 的分布密度 (参看本章习题 61). 显然，可改写 (21) 为

$$P(\eta\leqslant y|\xi=x)=\int_{-\infty}^{y} f(z|x)\,dx. \tag{23}$$

(丙) 在一般情形下，设 ξ,η 的联合分布函数为 $F(x,y)$. 我们定义 $P(\eta\leqslant y|\xi=x)$ 为在 $\xi=x$ 下， η 的条件分布函数为：

$$P(\eta\leqslant y|\xi=x)=\lim_{\alpha,\beta\to 0}\frac{F(x+\beta,y)-F(x-\alpha,y)}{F(x+\beta,\infty)-F(x-\alpha,\infty)},\ (\alpha>0,\beta>0). \tag{24}$$

只要右方极限存在. 注意，上式其实不过是 (14) 式当 C 逐渐缩小为单点集 $\{x\}$ 时的极限情形. 这个定义与 (19)(21) 是相容的. 实际上，如 $F(x,y)$ 有密度为 $f(x,y)$, 则 (24) 化为

$$P(\eta\leqslant y|\xi=x)=\lim_{\alpha,\beta\to 0}\frac{\int_{x-\alpha}^{x+\beta}\int_{-\infty}^{y} f(w,z)\,dzdw}{\int_{x-\alpha}^{x+\beta}\int_{-\infty}^{\infty} f(w,z)\,dzdw}.$$

以 $\beta+\alpha$ 除右方的分子，分母，取极限即得 (21) 几乎处处成立.

如果 ξ,η, 独立，则由 $F(x,y)=F_\xi(x)\cdot F_\eta(y)$ 及 (24) 得

$$P(\eta\leqslant y|\xi=x)=\lim_{\alpha,\beta\to 0}\frac{F_\xi(x+\beta)F_\eta(y)-F_\xi(x-\alpha)F_\eta(y)}{F_\xi(x+\beta)-F_\xi(x-\alpha)}=F_\eta(y). \tag{25}$$

这表示 η 在 $\xi=x$ 下的条件分布函数与 η 的 (无条件) 分布函数 $F_\eta(y)$ 一样. 因此, 在独立情况下, 对 ξ 的条件不影响 η 的分布与分布密度. 这个结论很是直观, 也很重要.

(丁) 以上定义不难推广到多维情况. 设 $\xi=(\xi_1,\cdots,\xi_n)$, $\eta=(\eta_1,\cdots,\eta_m)$ 分别为 n 及 m 维随机向量, 我们定义

$$\begin{aligned}&P(\eta_1\leqslant y_1,\cdots,\eta_m\leqslant y_m|\xi_1=x_1,\cdots,\xi_n=x_n)\\&=\lim_{C\to\{x\}}\frac{P(\eta_1\leqslant y_1,\cdots,\eta_m\leqslant y_m;\xi\in C)}{P(\xi\in C)},\end{aligned}\tag{26}$$

只要右方极限存在; 这里 $C\in\mathcal{B}_n$, $\{x\}$ 为含一个点 $(x_1,\cdots,x_n)$ 的集, "$C\to\{x\}$" 表示 $C\supset\{x\}$, 而且它的直径 $d=\sup\limits_{\substack{x\in C\\z\in C}}\|x-z\|\to 0$, 距离是 n 维欧氏空间的.

特别, 设 ξ,η,ζ 是连续型的随机变数, 有密度为 $f(x,y,z)$, 则类似于 (21)(22) 有

$$P(\zeta\leqslant z|\xi=x,\eta=y)=\frac{\int_{-\infty}^{z}f(x,y,w)\,dw}{\int_{-\infty}^{\infty}f(x,y,w)\,dw},\tag{27}$$

$$f(z|x,y)=\frac{f(x,y,z)}{\int_{-\infty}^{\infty}f(x,y,w)\,dw}.\tag{28}$$

§2.8 随机向量的变换

(一) 问题的一般提法. 随机向量经某变换后变为一新随机向量, 如何求后者的分布函数? 这问题无论在理论上或实际中都有重大的意义. 在叙述问题的准确提法以前, 先看一些例.

例 1 设随机变数 ξ 的分布函数为 $F_\xi(x)$, 作变换

$$\eta=a\xi+b,\ (a\neq 0).$$

以 $F_\eta(y)(y\in R_1)$ 表随机变数 η 的分布函数, 则

$$F_\eta(y)=P(a\xi+b\leqslant y)=P(a\xi\leqslant y-b).$$

如 $a>0$, 得

$$F_\eta(y)=P\Big(\xi\leqslant\frac{y-b}{a}\Big)\equiv F_\xi\Big(\frac{y-b}{a}\Big);$$

如 $a<0$, 得

$$F_\eta(y)=P\Big(\xi\geqslant\frac{y-b}{a}\Big)=1-P\Big(\xi<\frac{y-b}{a}\Big)=1-F_\xi\Big(\frac{y-b}{a}-0\Big).$$

若 $F_\xi(x)$ 有密度函数 $f_\xi(x)$, 显然 [1] $F_\eta(y)$ 有密度函数为

[1] 这时 $F_\xi(x)$ 连续, 故 $F_\xi(x-0)=F_\xi(x)$.

$$f_\eta(y) = \frac{1}{|a|} f_\xi\Big(\frac{y-b}{a}\Big). \qquad (1)$$

特别，如 ξ 有 $N(0,1)$ 分布，则 $\eta = \sigma\xi + a$ 有 $N(a,\sigma)$ 分布，$(\sigma > 0)$.

例 2 设 ξ 在 $[0,\pi]$ 中均匀分布，试求

$$\eta = \sin\xi$$

的分布函数 $F_\eta(y)$.

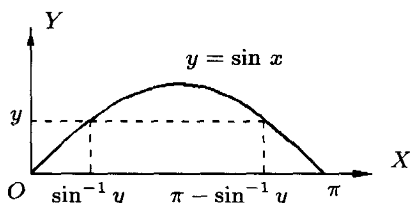


图 1

解 由于 $\xi \in [0,\pi]$, 故 η 只能取 $[0,1]$ 中的值. 对任意 $y \in [0,1]$, 由图 1 知

$$\begin{aligned} F_\eta(y) &= P(\eta \leqslant y) = P(\sin\xi \leqslant y) = P(0 \leqslant \xi \leqslant \sin^{-1} y) + P(\pi - \sin^{-1} y \leqslant \xi \leqslant \pi) \\ &= \frac{1}{\pi}[\sin^{-1} y + \pi - (\pi - \sin^{-1} y)] = \frac{2\sin^{-1} y}{\pi}. \end{aligned}$$

因此，

$$F_\eta(y) = \begin{cases} 0, & \text{如} y \leqslant 0, \\ \dfrac{2\sin^{-1} y}{\pi}, & \text{如} 0 \leqslant y \leqslant 1, \\ 1, & \text{如} y > 1, \end{cases} \qquad (2)$$

它的密度是

$$f(y) = \frac{2}{\pi\sqrt{1-y^2}}, \quad 0 \leqslant y \leqslant 1.$$

问题的一般提法如下：

设 n 维随机向量 $\xi = (\xi_1, \cdots, \xi_n)$ 的联合分布函数为 $F_\xi(x)$, $x = (x_1, \cdots, x_n) \in R_n$, 又设已给 m 个 n 元波雷耳可测函数 $u_i(x)$, $i = 1, \cdots, m$. 易见 [1] $\eta_i = u_i(\xi) = u_i(\xi_1, \cdots, \xi_n)$ 是一随机变数，因而 $\eta = (\eta_1, \cdots, \eta_m)$ 是一个 m 维随机向量，它自 ξ 经 $u_i(x)$ $i = 1, \cdots, m$ 变换得来. 我们的目的是求 η 的联合分布函数 $F_\eta(y)$, $y = (y_1, \cdots, y_m) \in R_m$.

理论上看，这问题并不困难. 实际上，对任意实数 $y_1, \cdots, y_m$, 考虑 R_n 中的子集

$$c_i = (x:\ u_i(x) \leqslant y_i) \in \mathcal{B}_n,\ C = \bigcap_{i=1}^{m} c_i \in \mathcal{B}_n.$$

则

$$\begin{aligned} F_\eta(y) &= P(\eta_1 \leqslant y_1, \cdots, \eta_m \leqslant y_m) = P(u_1(\xi) \leqslant y_1, \cdots, u_m(\xi) \leqslant y_m) \\ &= P(\xi \in c_1, \cdots, \xi \in c_m) = P(\xi \in C) = F_\xi(C), \end{aligned} \qquad (3)$$

其中 $F_\xi(A)(A \in \mathcal{B}_n)$ 表 ξ 的分布.

---

[1] 实际上，对任意 $a \in R_1$, 点 x 的集 $A = (x : u_i(x) \leqslant a) \in \mathcal{B}_n$, 故 ω− 集 $(\omega : u_i(\xi_1, \cdots, \xi_n) \leqslant a) = (\omega : (\xi_1, \cdots, \xi_n) \in A) \in \mathcal{F}$. 其次，如只需近似地求出 $F_\eta(y)$, 则可用下法. 仍以例 2 来说明. 利用 §6.1, 先作出 n 个独立，有相同均匀分布于 $[0,\pi]$ 中的随机变数 ξ_i, 并令 $\eta_i = \sin\xi_i$, $i = 1, \cdots, n$. 以 m 表满足 $\eta_i \leqslant y$ 的 i 的个数，则由 §3.3 中的强大数定理，当 n 充分大时，$\frac{m}{n}$ 近似地等于 $F_\eta(y)$. 对一般的 ξ 与 $\eta = u(\xi)$, 方法相同. 先利用 §6.2 以作出与 ξ 同分布的 n 个独立的 ξ_i, 令 $\eta_i = u(\xi_i)$, 又设 m 为满足 $\eta_i \leqslant y$ 的 i 的个数，则 $\frac{m}{n}$ 接近于 $F_\eta(y)$.

(3) 式表示，$F_\eta(y)$ 等于 ξ 落于 (即取值于) 集 C 中的概率. 如果 ξ 有联合密度为 $f_\xi(x_1,\cdots,x_n)$, 则 (3) 化为

$$F_\eta(y)=\int\cdots\int_C f_\xi(x_1,\cdots,x_n)\,dx_1\cdots dx_n. \tag{4}$$

然而在实际计算 $F_\xi(C)$ 时，却常常不容易，问题决定于两个因素：

1° 函数 $u_i(x)$ $i=1,\cdots,n$ 是简单，还是复杂？

2° ξ 的联合分布函数 $F_\xi(x)$ 是简单，还是复杂？

对 $u_i(x)$ 及 $F_\xi(x)$ 作相当强的假定 (如设 u_i 是 $x_1,\cdots,x_n$ 的线性函数，$F_\xi(x)$ 是独立随机变数的联合分布函数等等) 后，可以得到完满的结果.

(二) 一些重要的变换.

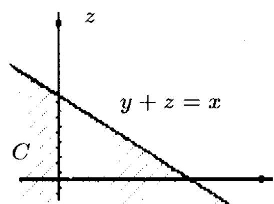


图 2

1° 和的分布. 设 $\xi=(\xi_1,\xi_2)$ 是二维随机向量，联合分布函数为 $F_\xi(x_1,x_2)$, 为方便计，改写为 $F(y,z)$. 试求随机变数 $Z=\xi_1+\xi_2$ 的分布函数 $G(x)$.

这时 $C=((y,z):\ y+z\leqslant x)$, 由 (3)

$$G(x)=F(C)=\iint_{y+z\leqslant x} dF(y,z).$$

这里及以后关于分布函数的积分都是勒贝格 - 司梯阶积分，它的定义及性质见参考书 [10] 或 [2]. 如 (ξ_1,ξ_2) 有联合密度 $f(y,z)$, 则 $G(x)=\iint\limits_{y+z\leqslant x} f(y,z)\,dydz=\int_{-\infty}^x\int_{-\infty}^\infty f(y,u-y)\,dydu$, 故知 Z 也有密度为

$$g(x)=G'(x)=\int_{-\infty}^{\infty} f(y,x-y)\,dy. \tag{4_1}$$

如 ξ_1,ξ_2 独立，分别有分布函数为 $F_1(y),F_2(z)$, 则由

$$F(y,z)=F_1(y)F_2(z)$$

得

$$\begin{aligned}G(x)&=\iint_{y+z\leqslant x} dF_1(y)dF_2(z)=\int_{-\infty}^{\infty}\left[F_2(z)\Big|_{-\infty}^{x-y}\right]dF_1(y)=\int_{-\infty}^{\infty}F_2(x-y)\,dF_1(y)\\&=\int_{-\infty}^{\infty}\left[F_1(y)\Big|_{-\infty}^{x-z}\right]dF_2(z)=\int_{-\infty}^{\infty}F_1(x-y)\,dF_2(y).\end{aligned} \tag{5}$$

一般地，称分布函数 (或分布)$G(x)$ 为二分布函数 (或分布)$F_1(x)$ 及 $F_2(x)$ 的卷积，如果对任意 $x\in R_1$, 有

$$G(x)=\int_{-\infty}^{\infty}F_2(x-y)\,dF_1(y)=\int_{-\infty}^{\infty}F_1(x-y)\,dF_2(y). \tag{6}$$

通常记 (6) 式为

$$G=F_2*F_1=F_1*F_2.$$

如 $F_1(x), F_2(x)$ 分别有密度为 $f_1(x), f_2(x)$, 则卷积 $G(x)$ 也有密度 $g(x)$ 为

$$g(x) = \int_{-\infty}^{\infty} f_2(x-y) f_1(y)\, dy = \int_{-\infty}^{\infty} f_1(x-z) f_2(z)\, dz. \tag{7}$$

由此可见：上述独立随机变数 ξ_1, ξ_2 的和 Z 的分布函数 $G(x)$ 是各自的分布函数的卷积；而且如 ξ_1, ξ_2 是连续型的，那么 Z 也是连续型的，三者的密度间有关系 (7).

(7) 式称为密度的卷积公式，并称其中的 $g(x)$ 为 $f_1(x), f_2(x)$ 的卷积.

如 ξ_1, ξ_2 是离散型的独立随机变数，ξ_1 只能取值 $x_i, (i=0,1,2,\cdots)$, ξ_2 只能取值 $y_j (j=0,1,2,\cdots)$, 那么 z 只能取值 $x_i + y_j (i,j=0,1,2,\cdots)$, 因而 z 也是离散型的，而且

$$\begin{aligned} P(Z=x) &= \sum_{i=0}^{\infty} P(\xi_2 = x - x_i) P(\xi_1 = x_i) \\ &= \sum_{j=0}^{\infty} P(\xi_1 = x - y_j) P(\xi_2 = y_j). \end{aligned} \tag{8}$$

这式与 (7) 式相当.

2° 差的分布. 差

$$Z = \xi_1 - \xi_2$$

的分布可类似求出；或者利用 $(\xi_1, -\xi_2)$ 的联合分布函数为

$$\begin{aligned} &P(\xi_1 \leqslant y, -\xi_2 \leqslant z) = P(\xi_1 \leqslant y, \xi_2 \geqslant -z) \\ &\quad = P(\xi_1 \leqslant y) - P(\xi_1 \leqslant y, \xi_2 < -z) = F_1(y) - F(y, -z-0) \end{aligned}$$

及 $\xi_1 - \xi_2 = \xi_1 + (-\xi_2)$ 而化为 1°, 详细的讨论从略.

3° 商的分布. 设 $\xi = (\xi_1, \xi_2)$ 是连续型二维随机向量，密度函数为 $f(y,z)$. 前面已经看到，ξ_i 也是连续型的，密度函数设为 $f_i(y), i=1,2$. 由于 $P(\xi_2 = 0) = 0$(见 §2.7(15)), 故可考虑随机变数

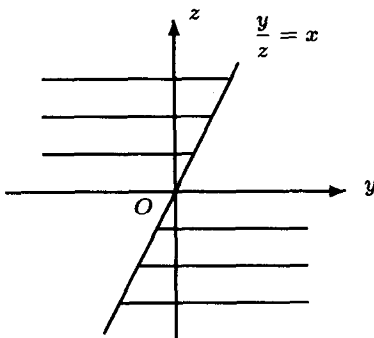


图 3

$$Z = \frac{\xi_1}{\xi_2},$$

试求 Z 的分布函数 $G(x)$.

$$\begin{aligned} G(x) &= P\left(\frac{\xi_1}{\xi_2} \leqslant x\right) = \iint\limits_{\frac{y}{z} \leqslant x} f(y,z)\, dydz \\ &= \int_0^{\infty} \int_{-\infty}^{xz} f(y,z)\, dydz + \int_{-\infty}^{0} \int_{xz}^{\infty} f(y,z)\, dydz. \end{aligned} \tag{9}$$

$G(x)$ 的密度 $g(x)$ 为

$$\begin{aligned} g(x) &= \int_0^{\infty} f(xz, z) z\, dz - \int_{-\infty}^{0} f(xz, z) z\, dz \\ &= \int_{-\infty}^{\infty} f(xz, z) |z|\, dz. \end{aligned} \tag{10}$$

如果 ξ_1,ξ_2 独立，则由 §2.7 系 2,(9),(10) 分别化为

$$G(x)=\int_0^\infty\int_{-\infty}^{xz}f_1(y)f_2(z)\,dydz+\int_{-\infty}^0\int_{xz}^\infty f_1(y)f_2(z)\,dydz$$

$$=\int_0^\infty F_1(xz)f_2(z)\,dz+\int_{-\infty}^0[1-F_1(xz)]f_2(z)\,dz; \tag{11}$$

$$g(x)=\int_{-\infty}^\infty f_1(xz)f_2(z)|z|\,dz. \tag{12}$$

(三) **连续型情况**. 现在来证明一个非常有用的定理，它适用于连续型的随机向量.

定理 1 设 n 维随机向量 $X=(X_1,\cdots,X_n)$ 的密度函数为 $f_X(x_1,\cdots,x_n)$, 又设 n 元函数 $u_i(x_1,\cdots,x_n)(i=1,\cdots,n)$ 满足条件

1′ 存在唯一反函数 $x_i(y_1,\cdots,y_n)$, 即存在方程

$$u_i(x_1,\cdots,x_n)=y_i \tag{13}$$

的唯一实值解 $x_i(y_1,\cdots,y_n)$, $(i=1,\cdots,n)$;

2′ $u_i(x_1,\cdots,x_n)$ 及 $x_i(y_1,\cdots,y_n)$ 都连续；

3′ 存在连续的偏导数 $\frac{\partial x_i}{\partial y_j},\frac{\partial u_i}{\partial x_j}$. 以 J 表雅各比行列式

$$J=\begin{vmatrix}\frac{\partial x_1}{\partial y_1}\cdots\frac{\partial x_1}{\partial y_n}\\ \vdots \\ \frac{\partial x_n}{\partial y_1}\cdots\frac{\partial x_n}{\partial y_n}\end{vmatrix},$$

则 n 维随机向量 $Y=(Y_1,\cdots,Y_n)$, 其中

$$Y_i=u_i(X_1,\cdots,X_n),\ (i=1,\cdots.n)$$

有密度函数为

$$f_Y(y_1,\cdots,y_n)=\begin{cases}f_X(x_1(y_1,\cdots,y_n),\cdots,x_n(y_1,\cdots,y_n))\cdot|J|,\\ \qquad\qquad\text{如 }(y_1,\cdots,y_n)\text{ 使 (13) 有解;}\\ 0,\qquad\qquad\text{反之.}\end{cases} \tag{14}$$

证 由 (3)

$$F_Y(y_1,\cdots,y_n)=F_X(C)=\int\cdots\int_C f_X(x_1,\cdots,x_n)dx_1\cdots dx_n.$$

作变换 $x_i=x_i(y_1,\cdots,y_n)$ 后，此积分化为

$$\int_{-\infty}^{y_1}\cdots\int_{-\infty}^{y_n}f_X(x_1(v_1,\cdots,v_n),\cdots,x_n(v_1,\cdots,v_n))|J|dv_1\cdots dv_n,$$

这得证 (14) 中第一结论. 其次，如对某 $y_1,\cdots,y_n$,(13) 无解，这表示 $Y_i=u_i(X_1,\cdots,X_n)$ 不能取 y_i 为值，故在此 $y_1,\cdots,y_n$ 上，Y 的密度应为 0. □

注 1. 如反函数不唯一，即如方程组 (13) 有多个解 $x_i^{(l)}(y_1,\cdots,y_n)$, $l=1,2,\cdots$, 则这时 $y-$ 空间中一个点，对应于 $x-$ 空间中多个点，将 $x-$ 空间分成若干部分，使 $y-$ 空间与每部分成一一对应，于是 Y 取值于 $y-$ 空间某集的概率就等于 X 取值于 $x-$ 空间中每部分里对应的集的概率的和. 故应对每一反函数运用 (14), 然后将所得结果相加，这样便求得 Y 的密度为:

$$f_Y(y_1,\cdots,y_n)=\begin{cases}\sum\limits_l f_X(x_1^{(l)}(y_1,\cdots,y_n),\cdots,x_n^{(l)}(y_1,\cdots,y_n))|J^{(l)}|, & \text{如 } (y_1,\cdots,y_n) \text{ 使 (13) 有解;}\\ 0, & \text{反之.}\end{cases} \tag{14$_1$}$$

下面的几个例子更清楚地说明了这一点.

注 2. 如果只给定 $m(<n)$ 个连续函数 $u_i(x_1,\cdots,x_n), i=1,\cdots,m$, 那么可补定义

$$u_j(x_1,\cdots,x_n)=x_j,\ (j=m+1,m+2,\cdots,n),$$

因而仍可利用上定理而得 $Y_n=(u_1(X_1,\cdots,X_n),\cdots,u_m(X_1,\cdots,X_n),X_{m+1},\cdots,X_n)$ 的密度 $f_{Y_n}(y_1,\cdots,y_m,y_{m+1},\cdots,y_n)$. 于是由 §2.6(三) 知 $Y_m=(u_1(X_1,\cdots,X_n),\cdots,u_m(X_1,\cdots,X_n))$ 的密度为

$$f_{Y_m}(y_1,\cdots,y_m)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{Y_n}(y_1,\cdots,y_m,y_{m+1},\cdots,y_n)dy_{m+1}\cdots dy_n.$$

下面的例对掌握定理 1 很有帮助.

例 3 设星球 A 至最近星球 B 的距离 X 的分布函数为

$$F_X(x)=P(X\leqslant x)=1-e^{-\frac{4}{3}\pi\lambda x^3},\ (x\geqslant 0),$$

(参看 §1.6 例 2), 试求 B 对 A 的引力 Y

$$Y=\frac{k}{X^2}\ (k\text{为正常数})$$

的密度函数 $f_Y(y)$.

解

$$f_X(x)=F_X'(x)=\begin{cases}4\pi\lambda x^2e^{-\frac{4}{3}\pi\lambda x^3}, & (x\geqslant 0),\\ 0, & (x<0).\end{cases}$$

设 $y>0$. 由 $y=\frac{k}{x^2}$ 得 $x=\pm\sqrt{\frac{k}{y}}$, 故有二反函数，对它们都有 $|J|=\left|\frac{dx}{dy}\right|=\frac{\sqrt{k}}{2}y^{-\frac{3}{2}}$. 由 (14$_1$)

$$f_Y(y)=\left[f_X\left(\sqrt{\frac{k}{y}}\right)+f_X\left(-\sqrt{\frac{k}{y}}\right)\right]\frac{\sqrt{k}}{2}y^{-3/2},$$

但 $f_X\left(-\sqrt{\frac{k}{y}}\right)=0$, 故

$$\begin{aligned}f_Y(y)&=f_X\left(\sqrt{\frac{k}{y}}\right)\frac{\sqrt{k}}{2}y^{-\frac{3}{2}}\\&=2\pi\lambda k^{\frac{3}{2}}y^{-\frac{5}{2}}\exp\left[-\frac{4}{3}\pi\lambda\left(\frac{k}{y}\right)^{\frac{3}{2}}\right],\ (y>0).\end{aligned}$$

如 $y \leqslant 0$, 则方程 $y = \frac{k}{x^2}$ 无解，故 $f_Y(y) = 0$.

例 4 设 X 的密度为 $f_X(x)$, 试求 $Y = \cos X$ 的密度.

解 1°. 设 $|y| \leqslant 1$, 则 $y = \cos x$ 的解为

$$x^{(l)}(y) = 2l\pi + \cos^{-1} y \text{ 及 } \overline{x}^{(l)}(y) = 2(l+1)\pi - \cos^{-1} y,$$

$l = 0, \pm 1, \pm 2, \cdots$, 共有无穷多个，$(\cos^{-1} y \leqslant \pi)$. 反函数的雅各比行列式的绝对值为

$$\left|\frac{dx^{(l)}(y)}{dy}\right| = \left|\frac{d\overline{x}^{(l)}(y)}{dy}\right| = \frac{1}{\sqrt{1-y^2}}.$$

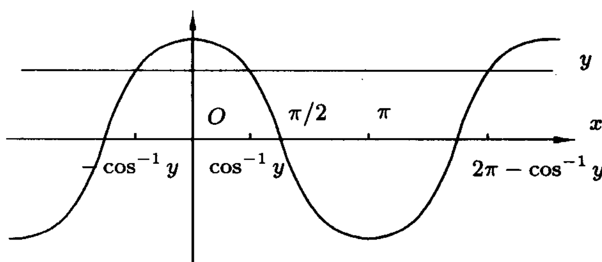


图 4

由 (14_1) 得

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}} \sum_{l=-\infty}^{\infty} [f_X(2l\pi + 2\pi - \cos^{-1} y) + f_X(2l\pi + \cos^{-1} y)], \quad (|y| \leqslant 1), \tag{14_2}$$

如 $|y| > 1$, 则方程 $y = \cos x$ 无解，故 $f_Y(y) = 0$.

2°. 自然，我们也可不用定理 1 而直接计算 Y 的分布函数 $F_Y(y)$. 对 $|y| \leqslant 1$,

$$F_Y(y) = \sum_{l=-\infty}^{\infty} P(\cos X \leqslant y, 2l\pi < X \leqslant 2(l+1)\pi),$$

由图 4 看出，$(\cos X \leqslant y, 2l\pi < X \leqslant 2(l+1)\pi)$ 重合于事件 $(2l\pi + \cos^{-1} y < X \leqslant 2(l+1)\pi - \cos^{-1} y)$, 因此

$$F_Y(y) = \sum_{l=-\infty}^{\infty} [F_X(2l\pi + 2\pi - \cos^{-1} y) - F_X(2l\pi + \cos^{-1} y)].$$

将此式对 y 微分，即得 (14_2).

3°. 至今我们对 $f_X(x)$ 未作任何假定. 现在考虑一特殊情形，设 X 在 $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ 中均匀分布，即

$$f_X(x) = \begin{cases} \dfrac{1}{\pi}, & |x| \leqslant \dfrac{\pi}{2}; \\ 0, & |x| > \dfrac{\pi}{2}. \end{cases}$$

则对 $0 \leqslant y \leqslant 1$,$(14_2)$ 的级数中除剩下两项外，其余都为 0, 故此时

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}}[f_X(\cos^{-1} y) + f_X(-\cos^{-1} y)] = \frac{2}{\pi\sqrt{1-y^2}},\ (0 \leqslant y \leqslant 1).$$

显然 $Y = \cos X$ 不能取 $(-\infty, 0)$ 及 $(1, \infty)$ 中的值，故当 y 属于此二区间时， $f_Y(y) = 0$.

例 5 $Y_1 = \sqrt{X_1^2 + X_2^2}$, $Y_2 = \frac{X_1}{X_2}$.

1° 试求 $Y = (Y_1, Y_2)$ 的密度 $f_Y(y_1, y_2)$, 假定 $X = (X_1, X_2)$ 有密度为 $f_X(x_1, x_2)$. 由

$$y_1 = \sqrt{x_1^2 + x_2^2},\ \ y_2 = \frac{x_1}{x_2}$$

解得二反函数

$$\begin{aligned} x_1^{(1)} &= \frac{y_1 y_2}{\sqrt{1+y_2^2}},\ \ x_2^{(1)} = \frac{y_1}{\sqrt{1+y_2^2}}; \\ x_1^{(2)} &= -x_1^{(1)},\ \ x_2^{(2)} = -x_2^{(1)}. \end{aligned}$$

容易算出

$$|J^{(1)}| = \begin{vmatrix} \frac{\partial x_1^{(1)}}{\partial y_1} & \frac{\partial x_1^{(1)}}{\partial y_2} \\ \frac{\partial x_2^{(1)}}{\partial y_1} & \frac{\partial x_2^{(1)}}{\partial y_2} \end{vmatrix} = \frac{|y_1|}{1+y_2^2},$$

同样， $|J^{(2)}| = \frac{|y_1|}{1+y_2^2}$ 于是由 (14_1)

$$f_Y(y_1, y_2) = \begin{cases} \frac{y_1}{1+y_2^2}\left[f_X\left(\frac{y_1 y_2}{\sqrt{1+y_2^2}}, \frac{y_1}{\sqrt{1+y_2^2}}\right) + f_X\left(\frac{-y_1 y_2}{\sqrt{1+y_2^2}}, \frac{-y_1}{\sqrt{1+y_2^2}}\right)\right], & \text{如} y_1 \geqslant 0; \\ 0, & \text{如} y_1 < 0. \end{cases}$$

2° 进一步，设

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma^2} e^{-\frac{x_1^2 + x_2^2}{2\sigma^2}},$$

即设 X_1, X_2 独立，有相同分布 $N(0, \sigma)$; 显然 $f_X(x_1, x_2) = f_X(-x_1, -x_2)$. 代入刚才所得结果，得

$$\begin{aligned} f_Y(y_1, y_2) &= \frac{y_1}{1+y_2^2}\frac{1}{\pi\sigma^2}\exp\left[-\left(\frac{y_1^2 y_2^2}{1+y_2^2} + \frac{y_1^2}{1+y_2^2}\right)\Big/2\sigma^2\right] \\ &= \frac{y_1}{\sigma^2}e^{-\frac{y_1^2}{2\sigma^2}} \cdot \frac{1}{\pi(1+y_2^2)},\ \ \text{如} y_1 \geqslant 0; \\ f_Y(y_1, y_2) &= 0,\ \ \text{如} y_1 < 0. \end{aligned}$$

由此我们还附带得到一个事先不易预见的结论：由于 $f_Y(y_1, y_2)$ 可表为 y_1 的函数 $\frac{y_1}{\sigma^2}e^{-\frac{y_1^2}{2\sigma^2}}$ 与 y_2 的函数 $\frac{1}{\pi(1+y_2^2)}$ 的乘积，可见 $Y_1 = \sqrt{X_1^2 + X_2^2}$ 与 $Y_2 = \frac{X_1}{X_2}$ 独立；而且 Y_1 的密度当 $y \geqslant 0$ 时是 $\frac{y}{\sigma^2}e^{-\frac{y^2}{2\sigma^2}}$, 当 $y < 0$ 时是 0,Y_2 的密度是 $\frac{1}{\pi(1+y^2)}$, $(-\infty < y < \infty)$. 这个结果所以难于预见，是因为 Y_1, Y_2 同是 X_1, X_2 的函数，因此，似乎 Y_1, Y_2 应该相关. 然而上述计算证明，这种表面印象是错误的 .[1]

---

[1] Y_1 的分布是瑞利分布， Y_2 的分布是柯西分布，见 94 页与 106 页.

例 6 积的分布. 设 $X=(X_1,X_2)$ 是连续型随机向量，乘积 X_1X_2 的分布及密度可仿求商的分布那样求出. 然而我们仍用定理 1 来求. 依照注 2, 令 $Y_1=X_1, Y_2=X_1X_2$. 由方程组 $y_1=x_1, y_2=x_1x_2$ 解得反函数 $x_1=y_1, x_2=\frac{y_2}{y_1}$; 又

$$J=\begin{vmatrix} 1 & 0 \\ -\frac{y_2}{y_1^2} & \frac{1}{y_1} \end{vmatrix}=\frac{1}{y_1},\quad |J|=\frac{1}{|y_1|}.$$

于是由 (14) 得

$$f_Y(y_1,y_2)=f_X\left(y_1,\frac{y_2}{y_1}\right)\frac{1}{|y_1|}.$$

由 §2.6(三) 得知 $Y_2=X_1X_2$ 的密度 $g(z)$ 及分布函数 $G(x)$ 分别为

$$g(z)=\int_{-\infty}^{\infty} f_X\left(y,\frac{z}{y}\right)\frac{1}{|y|}\,dy, \tag{15}$$

$$G(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_X\left(y,\frac{z}{y}\right)\frac{1}{|y|}\,dydz. \tag{16}$$

如 X_1,X_2 独立，分别有密度为 $f_1(x), f_2(x)$, 则上二式化为

$$g(z)=\int_{-\infty}^{\infty} f_1(y)f_2\left(\frac{z}{y}\right)\frac{1}{|y|}\,dy, \tag{17}$$

$$G(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_1(y)f_2\left(\frac{z}{y}\right)\frac{1}{|y|}\,dydz. \tag{18}$$

例 7 设独立随机变数 $\xi_1,\cdots,\xi_n$ 有相同的分布函数 $F(x)$, 它们的极大值是一新的随机变数 η:

$$\eta(\omega)=\max[\xi_1(\omega),\cdots,\xi_n(\omega)].$$

它在许多实际问题中起重要作用. 例如，设某河流第 i 年的最大径流量为 ξ_i, 则 n 年中的最大径流量为 η. 它的分布函数为 [1]

$$\begin{aligned} F_\eta(y) &= P(\eta\leqslant y)=P(\xi_1\leqslant y,\cdots,\xi_n\leqslant y) \\ &= \prod_{i=1}^{n} P(\xi_i\leqslant y)=[F(y)]^n; \end{aligned}$$

如 $F(y)$ 有密度函数为 $f(y)$, 则 $F_\eta(y)$ 也有密度为

$$f_\eta(y)=n[F(y)]^{n-1}f(y).$$

类似地，对极小值 ζ:

$$\zeta(\omega)=\min[\xi_1(\omega),\cdots,\xi_n(\omega)]$$

[1] 由此可见：如 $F(y)=y, (0\leqslant y\leqslant 1)$ 为均匀分布函数时， $F_\eta(y)=y^n=P(\xi_1\leqslant y^n)=P(\sqrt[n]{\xi_1}\leqslant y)$. 这表示 $\eta=\max(\xi_1,\cdots,\xi_n)$ 与 $\sqrt[n]{\xi_1}$ 同分布. 因而从同分布意义上看， η 与 $\sqrt[n]{\xi_1}$ 是等同的，故可用 η 来模拟 $\sqrt[n]{\xi_1}$(参看第六章). 此结论在蒙特卡罗计算方法中有用，因取极大值比开 n 次方易于实现. 极大值分布有许多应用，见 [8;41].

有

$$P(\zeta > y) = P(\xi_1 > y, \cdots, \xi_n > y) = \prod_{i=1}^{n} P(\xi_i > y) = [1 - F(y)]^n,$$

故

$$F_\zeta(y) = 1 - P(\zeta > y) = 1 - [1 - F(y)]^n.$$

(四) 三种分布的推导. 下列三种分布在数理统计中起着重要的作用.

定理 2 设相互独立的随机变数 $\xi_1, \cdots, \xi_n$ 有相同的正态分布 $N(a, \sigma)$, 考虑随机变数 χ^2,

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{\xi_i - a}{\sigma}\right)^2, \tag{19}$$

则随机变数 ζ

$$\zeta = \frac{\chi}{\sqrt{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\xi_i - a}{\sigma}\right)^2} \tag{20}$$

的密度函数为

$$g_n(x) = \begin{cases} 0 & \text{如} x < 0, \\ \dfrac{\sqrt{2\pi}}{\Gamma\left(\frac{n}{2}\right)}\left(\dfrac{x\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-\frac{nx^2}{2}}, & \text{如} x \geqslant 0. \end{cases} \tag{21}$$

证 令 $G_n(x) = P(\zeta \leqslant x)$, 因为 $\zeta \geqslant 0$, 故对 $x < 0$ 有 $G_n(x) = 0$. 下设 $x \geqslant 0$.

由于

$$G_n(x) = P\left(\frac{\chi}{\sqrt{n}} \leqslant x\right) = P(\chi^2 \leqslant nx^2), \tag{22}$$

故 $G_n(x)$ 等于随机向量 $(\frac{\xi_1 - a}{\sigma}, \cdots, \frac{\xi_n - a}{\sigma})$ 落于以原点为中心，以 $\sqrt{nx^2}$ 为半径的球

$$\sum_{i=1}^{n} x_i^2 \leqslant nx^2 \tag{23}$$

内的概率. 注意，$\frac{\xi_i - a}{\sigma}(i = 1, \cdots, n)$ 独立，有相同分布为 $N(0,1)$, 故它们的联合密度为

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}.$$

于是

$$G_n(x) = \int \cdots \int_{\sum_{i=1}^{n} x_i^2 \leqslant nx^2} \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} dx_1 \cdots dx_n. \tag{24}$$

下面的任务在于计算这 n 重积分. 作坐标变换

$$x_1 = \rho\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-1},$$
$$x_2 = \rho\cos\theta_1\cos\theta_2\cdots\sin\theta_{n-1},$$
$$\cdots\cdots$$
$$x_n = \rho\sin\theta_1$$

(当 $n=3$ 时 ρ,θ_1,θ_2 的几何意义见图 5)

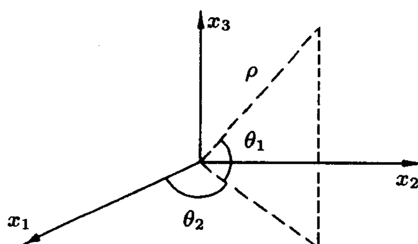


图 5

$$\begin{aligned}G_n(x) &= \int_{-\pi}^{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{x\sqrt{n}}\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\rho^2}{2}}\rho^{n-1}\cdot D(\theta_1,\cdots,\theta_{n-1})\,d\rho d\theta_1\cdots d\theta_{n-1}\\ &= c_n\int_0^{x\sqrt{n}} e^{-\frac{\rho^2}{2}}\rho^{n-1}\,d\rho,\end{aligned}\tag{25}$$

其中

$$c_n = \int_{-\pi}^{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left(\frac{1}{\sqrt{2\pi}}\right)^n D(\theta_1,\cdots,\theta_{n-1})\,d\theta_1\cdots d\theta_{n-1},$$

而

$$D(\theta_1,\cdots,\theta_{n-1}) = \frac{1}{\rho^{n-1}}\begin{vmatrix}\frac{\partial x_1}{\partial\rho} & \frac{\partial x_1}{\partial\theta_1} & \cdots & \frac{\partial x_1}{\partial\theta_{n-1}}\\ \frac{\partial x_2}{\partial\rho} & \frac{\partial x_2}{\partial\theta_1} & \cdots & \frac{\partial x_2}{\partial\theta_{n-1}}\\ \vdots & \vdots & \vdots & \\ \frac{\partial x_n}{\partial\rho} & \frac{\partial x_n}{\partial\theta_1} & \cdots & \frac{\partial x_n}{\partial\theta_{n-1}}\end{vmatrix},$$

右方行列式为变换行列式. 为求 c_n 于 (25) 中令 $x\to\infty$, 得

$$1 = \lim_{x\to\infty}G_n(x) = c_n\int_0^{\infty}e^{-\frac{\rho^2}{2}}\cdot\rho^{n-1}\,d\rho.$$

令 $t=\frac{\rho^2}{2}, \rho=\sqrt{2t}, d\rho=\frac{dt}{\sqrt{2t}}$, 由上式得

$$1 = c_n 2^{\frac{n}{2}-1}\int_0^{\infty}e^{-t}t^{\frac{n}{2}-1}\,dt = c_n 2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right),$$

其中 $\Gamma(\frac{n}{2})$ 为 Gamma 函数

$$\Gamma(p) = \int_0^{\infty}t^{p-1}e^{-t}\,dt,\quad (p>0)\tag{26}$$

在 $\frac{n}{2}$ 的值, 以 $c_n=\frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}$ 代入 (25), 即得

$$G_n(x) = \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}\int_0^{x\sqrt{n}}e^{-\frac{\rho^2}{2}}\cdot\rho^{n-1}\,d\rho,\ (x\geqslant 0).\tag{27}$$

微分后, 即知密度函数 $g_n(x)$ 由 (21) 式给出. □

当 $n=1$ 时，[1]

$$g_1(x)=\begin{cases}0, & x<0,\\ \sqrt{\dfrac{2}{\pi}}e^{-x^2/2}, & x\geqslant 0.\end{cases} \tag{28}$$

对非负的 x, $g_1(x)$ 等于 $N(0,1)$ 的密度在点 x 的值的二倍. 通常称它为反射正态分布，因如 ξ 有 $N(0,1)$ 分布，则 $|\xi|$ 的密度为 $g_1(x)$.

又当 $n=2,3$ 时，

$$g_2(x)=\begin{cases}0, & x<0,\\ 2xe^{-x^2}, & x\geqslant 0\end{cases};\qquad g_3(x)=\begin{cases}0, & x<0,\\ \dfrac{3\sqrt{6}}{\sqrt{\pi}}x^2e^{-(3/2)x^2}, & x\geqslant 0.\end{cases}$$

通常称密度函数为

$$f(x)=\begin{cases}0, & x<0,\\ \dfrac{x}{\sigma^2}e^{-x^2/2\sigma^2}, & x\geqslant 0,\end{cases}\quad (\sigma>0) \tag{29}$$

的分布为瑞利 (Rayleigh) 分布， $g_2(x)$ 是它的特殊情形 $\left(\sigma=\frac{1}{\sqrt{2}}\right)$. 称密度函数为

$$h(x)=\begin{cases}0, & x<0,\\ \dfrac{1}{\sigma^3}\sqrt{\dfrac{2}{\pi}}x^2e^{-x^2/2\sigma^2}, & x\geqslant 0,\end{cases}\quad (\sigma>0) \tag{29$_1$}$$

的分布为马克斯威尔 (Maxwell) 分布，它在统计物理中起着重要作用[2]. 上述 $g_3(x)$ 是它当 $\sigma=\frac{1}{\sqrt{3}}$ 的特殊情形.

系 1 (19) 中随机变数 χ^2 的密度函数为

$$f_n(x)=\begin{cases}0, & \text{如}x<0,\\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & \text{如}x\geqslant 0.\end{cases} \tag{30}$$

证

$$\begin{aligned}F_n(x)&\equiv P(\chi^2\leqslant x)=P\left(\frac{\chi}{\sqrt{n}}\leqslant\sqrt{\frac{x}{n}}\right)=G_n\left(\sqrt{\frac{x}{n}}\right)\\ &=\begin{cases}0, & (x<0),\\ \dfrac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}\displaystyle\int_0^{\sqrt{x}}e^{-\frac{\rho^2}{2}}\rho^{n-1}\,d\rho, & (x\geqslant 0).\end{cases}\end{aligned} \tag{31}$$

对 x 微分后即得 (30). □

[1]设 ξ 有 $N(0,\sigma)$ 分布，则 $|\xi|$ 有一般的反射正态分布，密度为 $\frac{g_1(x)}{\sigma}$, 而且其 k 阶矩：

$$E|\xi|^k=\sigma^k2^{\frac{k}{2}}\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}},$$ (参看 106 页).

[2]瑞利分布的数学期望与方差分别为 $\sqrt{\frac{\pi}{2}}\sigma$ 与 $\left(2-\frac{\pi}{2}\right)\sigma^2$. 马克斯威尔分布的数学期望与方差分别为 $2\sqrt{\frac{2}{\pi}}\sigma$ 与 $\left(3-\frac{8}{\pi}\right)\sigma^2$. 气体分子的速度 (绝对值) 服从马克斯威尔分布，它在统计物理中起重要作用.

称以 (30) 中 $f_n(x)$ 为密度的分布为具自由度 n 的 χ^2 分布，参数 n 称为自由度. 注意，虽然 (19) 中的 χ^2 依赖于 a,σ, 但它的密度却与 a,σ 无关，对不同的 $n, f_n(x)$ 图形见图 6.

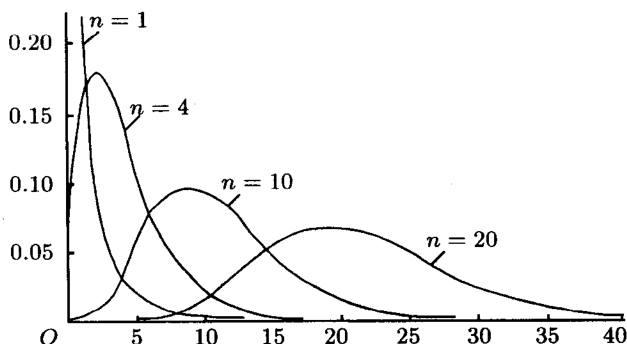


图 6 χ^2– 分布密度图. $n = 1, 4, 10, 20$

定理 3 设 ξ, z 为二独立随机变数，ξ 有 $N(0,1)$ 分布，z 有自由度为 n 的 χ^2– 分布，则

$$t = \xi \Big/ \sqrt{\frac{z}{n}} = \sqrt{n}\frac{\xi}{\sqrt{z}} \tag{32}$$

的密度为

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(\frac{x^2}{n}+1\right)^{-\frac{n+1}{2}}. \tag{33}$$

证 因 $\sqrt{z}/\sqrt{n}$ 与 (20) 中 $\frac{\chi}{\sqrt{n}}$ 同分布，故它的密度是 (21) 中的 $g_n(x)$; 又 ξ 的密度是 $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, 故由 (12) 得

$$\begin{aligned}
f_n(x) &= \int_0^\infty \frac{\sqrt{2n}}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{y\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-\frac{ny^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2y^2}{2}} y\,dy \\
&= \frac{1}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty \left(\frac{y\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-\frac{ny^2}{2}(1+\frac{x^2}{n})} ny\,dy.
\end{aligned}$$

作变换 $u = \frac{ny^2}{2}\left(\frac{x^2}{n}+1\right)$, 得

$$f_n(x) = \frac{\left(\frac{x^2}{n}+1\right)^{-\frac{n+1}{2}}}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty u^{\frac{n-1}{2}} e^{-u}\,du = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(\frac{x^2}{n}+1\right)^{-\frac{n+1}{2}}. \quad \square$$

称以 (33) 中的 $f_n(x)$ 为密度的分布为具自由度为 n 的 t– 分布 (或称学生分布), $f_n(x)$ 的图见 §2.5 常用分布表. 图 7 则将 $N(0,1)$ 正态分布的密度与 $n=2, n=5$ 的 t– 分布密度作了比较，当 n 很大时，这二者差别很小，实际上，只要 $n > 30$, 二者的值就相差无几了. 因此 $t(l), t(m)$ 的密度也非常接近，只要 l, m 充分大，譬如说，都大于 30. 这也可由 t– 分布表直接看出.

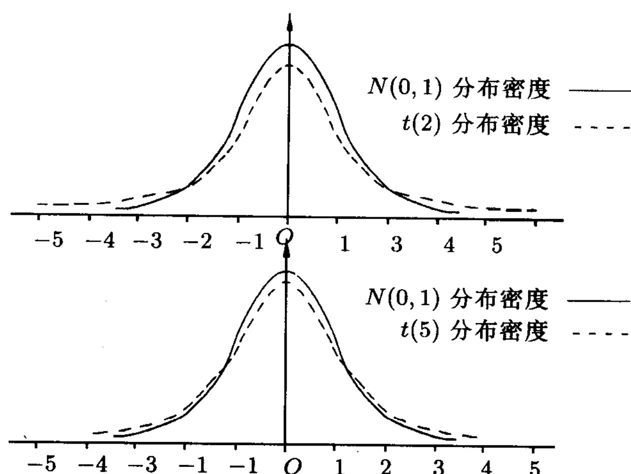


图 7 $n=2,5$ 时，t 分布密度与 $N(0,1)$ 正态分布密度的比较图

定理 4 设 χ^2_m,χ^2_n 为独立随机变数，分别具有自由度为 m 及 n 的 χ^2- 分布，令

$$\xi=\frac{\chi^2_m}{m},\quad \eta=\frac{\chi^2_n}{n},$$

则 $\zeta=\frac{\xi}{\eta}$ 的密度为

$$f(x)=\begin{cases}0, & \text{如}x\leqslant 0,\\ \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\dfrac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}, & \text{如}x>0.\end{cases}\tag{34}$$

证 由 (30) 推知 $\zeta=\frac{\chi^2_m}{m}$ 的分布密度为

$$f_\xi(y)=\begin{cases}0, & y\leqslant 0,\\ \dfrac{m^{\frac{m}{2}}}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}y^{\frac{m}{2}-1}e^{-\frac{my}{2}}, & y>0.\end{cases}$$

由 (12) 可见 $\zeta=\frac{\xi}{\eta}$ 的密度为

$$\begin{aligned}f(x)&=c\int_0^\infty y(xy)^{\frac{m}{2}-1}e^{-\frac{mxy}{2}}y^{\frac{n}{2}-1}e^{-\frac{ny}{2}}\,dy\\ &=cx^{\frac{m}{2}-1}\int_0^\infty y^{\frac{m+n}{2}-1}e^{-\frac{1}{2}y(mx+n)}\,dy,\end{aligned}$$

其中

$$c=\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}.\tag{35}$$

令 $z=y(mx+n)$, 则

$$\begin{aligned}f(x)&=c\frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}\int_0^\infty z^{\frac{m+n}{2}-1}e^{-\frac{z}{2}}\,dz\\ &=\frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}2^{\frac{m+n}{2}}\Gamma\left(\frac{m+n}{2}\right).\end{aligned}$$

由此及 (35) 即推出 (34). □

称以 (34) 中 $f(x)$ 为密度的分布为具有二自由度 m 及 n 的 $F-$ 分布，记为 $F_{m,n}$, 它的图形也见 §2.5 常用分布表.

§2.9 随机变数的数字特征

(一) **引进数字特征的原因**. 为了要完满的表达随机变数 ξ 的概率性质，自然应该知道 ξ 的分布函数 $F(x)$, 然而在许多实际问题中，并不需要知道 ξ 的一切概率性质，而只需了解它的某一性质就够了. 例如，以 ξ 表普阿松流在时间 $(0,1)$ 中出现的质点数，有时只要知道平均的质点数，即 ξ 的平均数，而平均数是下面要讲的数字特征之一. 引进数字特征的另一原因是：有时很难精确地求出 $F(x)$, 因之不得不退而求其次.

数字特征一般被 $F(x)$ 确定；它是对 $F(x)$ 进行某种运算的结果. 因此，我们可以引进任意多个数字特征，只须引进任意多种运算，问题在于哪些数字特征才有实际的或理论上的意义. 一般地，重要的是反映随机变数 ξ 下列方面的性质的数字特征：

1. ξ 的集中位置：注意 $\xi=\xi(\omega)$ 是 ω 的函数，对应于不同的 $\omega, \xi(\omega)$ 的值可能不同，因而希望知道大多数的 $\xi(\omega)$ 集中在那里，能够粗略地 (未必精确地) 满足这一要求的是 ξ 的平均值，也叫 ξ 的数学期望；另一数字特征是中数，它们的严格定义留待以后.

2. ξ 的集中程度：除了知道集中的位置外，有时还要知道集中的程度，是高度的集中还是高度的分散？反映这一点的数字特征是方差和内极差等 [1].

3. 两个随机变数的相关程度：设有两个随机变数 ξ 与 η, 希望知道它们的概率联系是否紧密，反映相关程度的有相关系数.

当然，还有一些重要的其它的数字特征如矩，半不变量等.

直观的想法大致如此.

(二) **数学期望**. 如上所述，我们希望引进一个数字特征，来反映 ξ 的值的集中位置. 这使我们联想到力学中重心的概念，因为重心是反映力学系统的集中位置的. 设有一个一维力学系统 S, 它由 n 个质点构成，第 i 个质点的坐标为 a_i, 在 a_i 点上的质量为 $p_i, (i=1,\cdots,n)$, 那么力学中定义 S 的重心的坐标为

$$\sum_{i=1}^{n} a_i p_i \Big/ \sum_{i=1}^{n} p_i. \tag{1}$$

受重心定义的启发，对密度矩阵为 $\begin{pmatrix} a_0 & a_1 & \cdots \\ p_0 & p_1 & \cdots \end{pmatrix}$ 的离散型随机变数 ξ, 我们定义 ξ 的数学期望 $E\xi$ 为

$$E\xi=\sum_{i=0}^{\infty} a_i p_i, \tag{2}$$

(注意 $\sum\limits_{i=0}^{\infty} p_i=1$), 只要右方级数绝对收敛.

收敛的绝对性保证 $\sum\limits_{i=0}^{\infty} a_i p_i$ 的值不随级数项排列的次序改变而改变，这是必需的，因为数学期望反映客观实在 (ξ 的集中位置), 它自然不应随排序的主观性而改变. 收敛的绝

[1]此外，为了克服方差不一定存在的缺点，还有所谓散布度，也是反映分布的集中程度的数字特征，见 [18],§11.

对性还可以使我们充分运用勒贝格积分这一有力工具.

特别，如 ξ 的密度矩阵为 $\binom{a_1 \ \cdots \ a_n}{\frac{1}{n} \ \cdots \ \frac{1}{n}}$(古典型), 那么 $E\xi = \frac{1}{n}\sum\limits_{i=1}^{n} a_i$ 化为 $a_1, \cdots, a_n$ 的平均值. 所以数学期望是平均值的推广，即"加权"平均值.

与 (2) 对应，对密度函数为 $f(x)$ 的连续型随机变数 ξ, 如果 $\int_{-\infty}^{\infty}|x|f(x)\,dx < \infty$, 我们定义

$$E\xi = \int_{-\infty}^{\infty} xf(x)\,dx, \tag{3}$$

右方是勒贝格积分.

对一般的随机变数 ξ, 有分布函数 $F(x)$, 如果 $\int_{-\infty}^{\infty}|x|\,dF(x) < \infty$, 我们定义 ξ 的数学期望 $E\xi$ 为

$$E\xi = \int_{-\infty}^{\infty} x\,dF(x), \tag{4}$$

右方是勒贝格 - 司梯阶积分.

这定义还需要稍许推广. 设 $g(x), x \in R_1$ 是一元波雷耳可测函数，在 §2.7 系 1 中已知 $g(\xi)$ 是一随机变数，仍以 $F(x)$ 表 ξ 的分布函数. 如果勒贝格 - 司梯阶积分 $\int_{-\infty}^{\infty} g(x)\,dF(x)$ 绝对收敛，即如 $\int_{-\infty}^{\infty}|g(x)|\,dF(x) < \infty$, 就定义 $g(\xi)$ 的数学期望 (亦称平均值)$Eg(\xi)$ 为

$$Eg(\xi) = \int_{-\infty}^{\infty} g(x)\,dF(x). \tag{5}$$

由于 $|Eg(\xi)| \leqslant \int_{-\infty}^{\infty}|g(x)|\,dF(x)$, 可见如 $g(\xi)$ 有数学期望，则 $Eg(\xi)$ 必有穷 [1]. 对连续型的 $F(x)$, 如 $F'(x) = f(x)$, 则 (5) 化为

$$Eg(\xi) = \int_{-\infty}^{\infty} g(x)f(x)\,dx; \tag{5_1}$$

对离散型的 $F(x)$, 如它的密度矩阵是 $\binom{a_0 \ a_1 \ \cdots}{p_0 \ p_1 \ \cdots}$, 则 (5) 化为

$$Eg(\xi) = \sum_{i=0}^{\infty} g(a_i)p_i. \tag{5_2}$$

现在将 $g(x)$ 特殊化，就得到各种数字特征.

a. 令 $g(x) = x^k, (k \geqslant 0)$, 称 $E\xi^k$ 为 ξ 的 k 阶矩，记为 m_k.

b. 令 $g(x) = |x|^k, (k \geqslant 0)$, 称 $E|\xi|^k$ 为 ξ 的 k 阶绝对矩.

c. 令 $g(x) = (x - E\xi)^2$. 称 $E(\xi - E\xi)^2$ 为 ξ 的方差，记为 $D\xi$, 即

$$D\xi = \int_{-\infty}^{\infty} (x - E\xi)^2\,dF(x). \tag{6}$$

当 $F(x)$ 是连续型或离散型时，上式分别化为

$$D\xi = \int_{-\infty}^{\infty} (x - E\xi)^2 f(x)\,dx, \tag{6_1}$$

$$D\xi = \sum_{i=0}^{\infty} (a_i - E\xi)^2 p_i. \tag{6_2}$$

---

[1] 以后"$g(\xi)$ 有数学期望","$g(\xi)$ 的数学期望存在"与"$g(\xi)$ 的数学期望有穷"三句话是同一意思.

方差是表示 ξ 对它的平均值 (即数学期望)$E\xi$ 的分散程度的一个指标. $D\xi$ 越大，某些 $(a_i - E\xi)^2 p_i$ 也越大，即分散程度大；反之亦然. 特别，如 $D\xi = 0$, 则 $a_i \equiv E\xi$(一切 i), 这表示 ξ 有单点分布 $\binom{E\xi}{1}$, 这是高度集中的特殊情形. 这里我们用 (6_2) 来解释是为了便于说明，其实这事实对一般的 $F(x)$ 都正确. (6) 中采用平方是为了保证一切离差都起正面的作用，因而都被考虑在 $D\xi$ 中. 否则，如用一次方，正，负离差就有相互抵消的可能. 通常可用下式来计算方差. 由 (6)

$$\begin{aligned} D\xi &= \int_{-\infty}^{\infty} x^2\, dF(x) - 2E\xi \int_{-\infty}^{\infty} x\, dF(x) + (E\xi)^2 \\ &= E\xi^2 - (E\xi)^2. \end{aligned} \tag{7}$$

d. 令 $g(x) = (x - E\xi)^k, (k \geqslant 0)$. 称 $E(\xi - E\xi)^k$ 为 ξ 的 k 阶中心矩，记为 c_k.

e. 令 $g(x) = |x - E\xi|^k, (k \geqslant 0)$. 称 $E|\xi - E\xi|^k$ 为 ξ 的 k 阶绝对中心矩.

中心矩可通过矩来表达：对正整数 k, 有

$$\begin{aligned} c_k = E(\xi - E\xi)^k &= \sum_{i=0}^{k} \binom{k}{i} (-E\xi)^{k-i} \cdot E\xi^i \\ &= \sum_{i=0}^{k} \binom{k}{i} (-m_1)^{k-i} m_i. \end{aligned}$$

例如:

$$\left.\begin{array}{l} c_0 = 1, \quad c_1 = 0, \\ c_2 = m_2 - m_1^2, \quad c_3 = m_3 - 3m_2m_1 + 2m_1^2, \\ c_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4, \ \cdots \end{array}\right\}. \tag{8}$$

矩也可通过中心矩来表达:

$$\begin{aligned} m_k = E\xi^k &= E[(\xi - m_1) + m_1]^k \\ &= \sum_{i=0}^{k} \binom{k}{i} E(\xi - m_1)^{k-i} m_1^i \\ &= \sum_{i=0}^{k} \binom{k}{i} c_{k-i} m_1^i. \end{aligned}$$

现在来考虑多变量情形. 设 $\xi_1(\omega), \cdots, \xi_n(\omega)$ 是 n 个随机变数， $g(x_1, \cdots, x_n)$ 是 n 元波雷耳可测函数， $g(\xi_1(\omega), \cdots, \xi_n(\omega))$ 是随机变数，以 $F(x_1, \cdots, x_n)$ 表 $(\xi_1(\omega), \cdots, \xi_n(\omega))$ 的分布函数，完全类似地，定义 $g(\xi_1(\omega), \cdots, \xi_n(\omega))$ 的数学期望为

$$Eg(\xi_1, \cdots, \xi_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \cdots, x_n) dF(x_1, \cdots, x_n), \tag{9}$$

只要右方的勒贝格司梯阶积分绝对收敛.

(三) **一些数字特征的性质**. 以下总设 $\xi, \xi_1, \cdots, \xi_n$ 都是随机变量，而且所用到的数字特征存在.

定理 1 数学期望有下列性质:

(i) 如 c 是常数，则 $Ec = c$.

(ii) 线性：对任意常数 $c_i, i = 1, \cdots, n$, 有

$$E\left(\sum_{i=1}^{n} c_i \xi_i\right) = \sum_{i=1}^{n} c_i E\xi_i. \tag{10}$$

(iii) 如 $\xi_1, \cdots, \xi_n$ 独立，则

$$E\xi_1\xi_2\cdots\xi_n = \prod_{i=1}^{n} E\xi_i. \tag{11}$$

(iv) 函数 $f(x) = E(\xi - x)^2$ 当 $x = E\xi$ 时取最小值 $D\xi$.

证 常数 c 也是随机变数，这是因为把它看成恒等于 c 的函数时，§2.1(4) 式显然满足. 它是离散型的，密度矩阵为 $\binom{c}{1}$, 由 (2) 得

$$Ec = c \cdot 1 = c.$$

在 (9) 中取 $g(x_1, \cdots, x_n) = \sum\limits_{i=1}^{n} c_i x_i$, 并以 $F_i(x)$ 表 ξ_i 的分布函数，得 [1]

$$\begin{aligned}
E\left(\sum_{i=1}^{n} c_i \xi_i\right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (c_1 x_1 + \cdots + c_n x_n)\, dF(x_1, \cdots, x_n) \\
&= \sum_{i=1}^{n} c_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i\, dF(x_1, \cdots, x_n) = \sum_{i=1}^{n} c_i \int_{-\infty}^{\infty} x_i\, dF_i(x_i) \\
&= \sum_{i=1}^{n} c_i E\xi_i
\end{aligned} \tag{12}$$

今证 (iii): 由 $\xi_1, \cdots, \xi_n$ 的独立性得

$$F(x_1, \cdots, x_n) = \prod_{i=1}^{n} F_i(x_i).$$

---

[1] (12) 式中第三个等号的成立可由积分的性质推出：例如，当 $n = 2$ 时

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1\, dF(x_1, x_2) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{2^n} F\left(\left[\frac{k-1}{2^n} < x_1 \leqslant \frac{k}{2^n}\right] \times R_1\right) \\
&= \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{2^n} \cdot \left[F\left(\frac{k}{2^n}, \infty\right) - F\left(\frac{k-1}{2^n}, \infty\right)\right] \\
&= \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{2^n} F_1\left(\left[\frac{k-1}{2^n} < x_1 \leqslant \frac{k}{2^n}\right]\right) = \int_{-\infty}^{\infty} x_1\, dF_1(x_1),
\end{aligned}$$

其中 R_1 表一维空间. 对一般的 n 证明完全类似.

因而由实变函数论中富比尼 (Fubini) 定理 [1] 得

$$\begin{aligned}E\xi_1\cdots\xi_n &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1\cdots x_n\,dF(x_1,\cdots,x_n)\\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1\cdots x_n\,dF_1(x_1)\cdots dF_n(x_n)\\ &= \prod_{i=1}^{n}\int_{-\infty}^{\infty} x_i\,dF_i(x_i) = \prod_{i=1}^{n} E\xi_i.\end{aligned}$$

为证 (iv), 注意由 (i),(ii) 得

$$f(x) = E\xi^2 - 2xE\xi + x^2,\quad \frac{df(x)}{dx} = 2x - 2E\xi. \tag{13}$$

$\frac{df(x)}{dx} = 0$ 的根为 $E\xi$, 既然 $\frac{d^2f(x)}{dx^2} = 2 > 0$, 故 $f(x)$ 在 $x = E\xi$ 取最小值 $E(\xi - E\xi)^2 = D\xi$. □

这里的 (iv) 是可以把 $E\xi$ 看成 ξ 的集中位置的另一理由.

定理 2 方差具有下列性质：设 c, c_i 为常数：

(i) $Dc = 0$.

(ii) $D\left(\sum\limits_{i=1}^{n} c_i\xi_i\right) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} c_ic_jE(\xi_i - E\xi_i)(\xi_j - E\xi_j)$.

(iii) 如 $\xi_1,\cdots,\xi_n$ 独立，则

$$D\left(\sum_{i=1}^{n} c_i\xi_i\right) = \sum_{i=1}^{n} c_i^2 D\xi_i. \tag{14}$$

证 由 (7) 得

$$Dc = Ec^2 - (Ec)^2 = c^2 - c^2 = 0 \tag{15}$$

$$\begin{aligned}D\left(\sum_{i=1}^{n} c_i\xi_i\right) &= E\left[\sum_{i=1}^{n} c_i(\xi_i - E\xi_i)\right]^2 = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j(\xi_i - E\xi_i)(\xi_j - E\xi_j)\right]\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jE(\xi_i - E\xi_i)(\xi_j - E\xi_j).\end{aligned} \tag{16}$$

如 $\xi_1,\cdots,\xi_n$ 独立，则由定理 1

$$E\xi_i\xi_j = E\xi_iE\xi_j,\ (i,j = 1,\cdots,n,\ i \neq j), \tag{17}$$

[1] 参看 [10] 第五章 §9. 作为定理 1 性质 (ii) 的一个应用，考虑 §1.4 中的例 7, 即 n 张信纸与信封的配对问题. 设配好对的对数为 ξ, 试求 $E\xi$. 采用那里的记号，按定义得 $E\xi = \sum\limits_{r=0}^{n} rq_r = \sum\limits_{r=1}^{n}\frac{r}{r!}\sum\limits_{k=0}^{n-r}\frac{(-1)^k}{k!}$. 但另方面，也可利用 (ii) 来计算：定义 $\xi_k = 1$ 或 0, 视第 k 号信纸是或否装入同号码的信封而定. 于是，$\xi = \xi_1 + \cdots + \xi_n$. 故 $E\xi = E\xi_1 + \cdots + E\xi_n$. 但由 §1.4 例 7,$E\xi_k = 1\cdot P(\xi_k = 1) + 0\cdot P(\xi_k = 0) = P(\xi_k = 1) = \frac{1}{n}$. 从而 $E\xi = n\cdot\frac{1}{n} = 1$. 这样一来，我们便用概率的方法证明了下列恒等式 $\sum\limits_{r=1}^{n}\frac{1}{(r-1)!}\sum\limits_{k=0}^{n-r}\frac{(-1)^k}{k!} = 1$.

从而 $E(\xi_i - E\xi_i)(\xi_j - E\xi_j) = 0$, 故 (16) 化为

$$D\left(\sum_{i=1}^{n} c_i\xi_i\right) = \sum_{i=1}^{n} c_i^2 D\xi_i. \quad \square \tag{18}$$

特别注意下式中的加号，如 ξ_1, ξ_2 独立，由 (14) 有

$$D(\xi_1 - \xi_2) = D\xi_1 + D\xi_2. \tag{19}$$

顺便指出，由上面证明可见，如果 (17) 成立，那么除去独立性的假定后，(14) 仍然成立.

现在来研究 k 阶绝对矩 $a_k = E|\xi|^k$, $k = 0, 1, 2, \cdots$, 由于

$$|\xi|^k \leqslant 1 + |\xi|^{k+1}, \quad a_k \leqslant 1 + a_{k+1},$$

可见如高阶绝对矩有穷，则低阶绝对矩也有穷；故如高阶矩 m_n 存在，则低阶矩 $m_1, \cdots, m_{n-1}$ 也存在.

***定理 3** 如 $a_n < \infty$, 则

$$\sqrt[k]{a_k} \leqslant \sqrt[k+1]{a_{k+1}}, \quad k = 1, 2, \cdots, n-1. \tag{20}$$

证 考虑 u, v 的函数

$$a_{k-1}u^2 + 2a_k uv + a_{k+1}v^2 = \int_{-\infty}^{\infty} \left[u|x|^{\frac{k-1}{2}} + v|x|^{\frac{k+1}{2}}\right]^2 dF(x), \tag{21}$$

由右方可见这函数对任意数 u, v 非负，因此必定

$$a_k^2 - a_{k-1}a_{k+1} \leqslant 0, \quad 即 a_k^2 \leqslant a_{k-1}a_{k+1},$$
$$a_k^{2k} \leqslant a_{k-1}^k a_{k+1}^k.$$

令 $k = 1, 2, \cdots, n-1$, 得

$$a_1^2 \leqslant a_0 a_2, \quad a_2^4 \leqslant a_1^2 a_3^2, \quad \cdots, \quad a_{n-1}^{2(n-1)} \leqslant a_{n-2}^{n-1} a_n^{n-1},$$

其中 $a_0 = 1$. 将这些式中前 k 个的左，右方分别相乘，化简即得 $a_k^{k+1} \leqslant a_{k+1}^k$, 开 $k(k+1)$ 次方即得 (20). □

如果对 ξ, $E\xi, D\xi$ 都存在，有时考虑

$$\xi^* = \frac{\xi - E\xi}{\sqrt{D\xi}} \quad (如 D\xi > 0) \tag{22}$$

更方便，称 ξ^* 为 ξ 的标准化随机变数，显然

$$E\xi^* = 0, \quad D\xi^* = 1. \tag{23}$$

(四) 条件数学期望. 设事件 B 使 $P(B) > 0$, 令

$$F(x|B) = \frac{P(\xi \leqslant x, B)}{P(B)},$$

它是在事件 B 下 ξ 的条件分布函数. 如果积分 $\int_{-\infty}^{\infty} x\,dF(x|B)$ 绝对收敛，称此积分为 ξ 在事件 B 下的条件数学期望，并记为 $E(\xi|B)$, 显然. $E\xi = E(\xi|\Omega)$.

设事件 $B_1, B_2, \cdots, B_n$ 互不相交，$P(B_i) > 0, (i = 1, 2, \cdots, n)$ 而且 $\bigcup_{i=1}^{n} B_i = \Omega$, 由全概率公式

$$F(x) = P(\xi \leqslant x) = \sum_{i=1}^{n} P(\xi \leqslant x|B_i)P(B_i) = \sum_{i=1}^{n} F(x|B_i)P(B_i).$$

故得

$$E\xi = \int_{-\infty}^{\infty} x\,dF(x) = \sum_{i=1}^{n} E(\xi|B_i)P(B_i). \tag{24}$$

此式称为全数学期望公式.

正如全概率公式有利于计算事件的概率一样，(24) 式也有助于计算 $E\xi$.

如 ξ 与 B 独立，就是说，如对任意实数 x, 有

$$P([\xi \leqslant x] \cap B) = P(\xi \leqslant x)P(B).$$

那么由 (24), 显然 $F(x|B) = F(x)$, 于是

$$E(\xi|B) = \int_{-\infty}^{\infty} x\,dF(x|B) = \int_{-\infty}^{\infty} x\,dF(x) = E\xi.$$

现在考虑两个随机变数 ξ 与 η 的情形，在 §2.7(二) 中已定义过在 $\eta = y$ 的条件下，ξ 的条件分布函数 $P(\xi \leqslant x|\eta = y)$ 以及条件密度 $f(x|y)$; 简记 $P(\xi \leqslant x|\eta = y)$ 为 $F(x|y)$. 利用它们，就可定义：在 $\eta = y$ 下，$g(\xi)$ 的条件数学期望 $E(g(\xi)|y)$ 为

$$E(g(\xi)|y) = \int_{-\infty}^{\infty} g(x)\,dF(x|y), \tag{25}$$

只要右方积分存在；这里 $g(x)$ 为波雷耳可测函数. 如果条件密度存在，就有

$$E(g(\xi)|y) = \int_{-\infty}^{\infty} g(x)f(x|y)\,dx.$$

当 y 固定时，$E(g(\xi)|y)$ 是一个常数. 现在我们换一个观点，把 y 看成自变数，那么 $E(g(\xi)|y)$ 是 y 的函数，而且由 §2.7(22), 一般地它还是 y 的波雷耳可测函数. 以 η 代替 y 后，作为随机变数 η 的函数，$E(g(\xi)|\eta)$ 也是一个随机变数；它的数学期望是

$$\begin{aligned} E[E(g(\xi)|\eta)] &= \int_{-\infty}^{\infty} E(g(\xi)|y)f_\eta(y)\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)f(x|y)\,dx\,f_\eta(y)\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)f(x,y)\,dxdy = \int_{-\infty}^{\infty} g(x)f_\xi(x)\,dx = Eg(\xi), \end{aligned}$$

其中 $f(x,y)$ 是 (ξ, η) 的联合密度. 公式

$$Eg(\xi) = E[E(g(\xi)|\eta)] \tag{26}$$

是 (24) 的类似式，它表明：条件期望的平均值等于 (无条件) 平均值.

如果 ξ 与 η 相互独立，则因 $F(x|y)=F_\xi(x)$, 由 (25)

$$E(g(\xi)|y)=\int_{-\infty}^{\infty} g(x)\,dF_\xi(x)=Eg(\xi),$$

故此时条件期望与 (无条件) 期望一致.

(五) **例.**

例 1 设 ξ 有二项分布：

$$p_k=P(\xi=k)=\binom{n}{k}p^kq^{n-k},\quad k=0,1,\cdots,n.$$

则有

$$\begin{aligned}E\xi&=\sum_{k=0}^{n}kp_k=\sum_{k=1}^{n}k\frac{n!}{k!(n-k)!}p^kq^{n-k}\\&=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}\\&=np\sum_{k=0}^{n-1}\frac{(n-1)!}{k!(n-1-k)!}p^kq^{n-1-k}\\&=np[p+q]^{n-1}=np.\end{aligned}$$

$$\begin{aligned}D\xi=E\xi^2-(E\xi)^2&=\sum_{k=0}^{n}k^2\frac{n!}{k!(n-k)!}p^kq^{n-k}-(np)^2\\&=np\sum_{k=1}^{n}k\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}-(np)^2\\&=np\left[\sum_{k=0}^{n-1}k\frac{(n-1)!}{k!(n-1-k)!}p^kq^{n-1-k}+(p+q)^{n-1}\right]-(np)^2\\&=np[(n-1)p+1]-(np)^2=np(np+q)-(np)^2=npq.\end{aligned}$$

由于二项分布被二参数 n,p 所决定，由上可见，这二参数又由 $E\xi, D\xi$ 所决定，故二项分布完全被它的数学期望与方差所决定. 下面的例 2, 例 3 表明，这一点对普阿松分布及正态分布也正确.

例 2 设 ξ 有普阿松分布：

$$p_k=P(\xi=k)=e^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\cdots.$$

$$E\xi=\sum_{k=0}^{\infty}kp_k=\sum_{k=0}^{\infty}ke^{-\lambda}\frac{\lambda^k}{k!}=\lambda\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=\lambda e^{-\lambda}e^{\lambda}=\lambda.$$

$$
\begin{aligned}
D\xi &= E\xi^2 - (E\xi)^2 = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} - \lambda^2 \\
&= \lambda \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} - \lambda^2 \\
&= \lambda \sum_{k=1}^{\infty} (k-1) e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} + \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned}
$$

例 3 设 ξ 有正态分布 $N(a,\sigma)$, 则

$$
\begin{aligned}
E\xi &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + a) e^{-\frac{z^2}{2}}\, dz \quad \left(z = \frac{x-a}{\sigma}\right) \\
&= \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\, dz = a. \qquad (27) \\
D\xi &= E(\xi - a)^2 = \int_{-\infty}^{\infty} (x-a)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}}\, dz \quad \left(z = \frac{x-a}{\sigma}\right) \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[- z e^{-\frac{z^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\, dz \right] \quad (\text{分部积分}) \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \sigma^2. \qquad (28)
\end{aligned}
$$

因此, ξ 的标准化随机变数为 $\xi^* = \frac{\xi - a}{\sigma}$.

$$
c_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^k e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx = \frac{\sigma^k}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-\frac{x^2}{2}}\, dx.
$$

当 k 为奇数时, 被积函数是奇函数, 故

$$
c_k = 0;
$$

当 k 为偶数时, 作变换 $x^2 = 2z$ 得

$$
\begin{aligned}
c_k &= \sqrt{\frac{2}{\pi}} \sigma^k \int_0^{\infty} x^k e^{-\frac{x^2}{2}} dx \\
&= \sqrt{\frac{2}{\pi}} \sigma^k 2^{\frac{k-1}{2}} \int_0^{\infty} z^{\frac{k-1}{2}} e^{-z}\, dz \\
&= \sqrt{\frac{2}{\pi}} \sigma^k 2^{\frac{k-1}{2}} \Gamma\left(\frac{k+1}{2}\right) = \sigma^k (k-1)(k-3)\cdots 1.
\end{aligned}
$$

试求 k 阶绝对中心矩 $a_k = E|\xi - E\xi|^k$. 当 k 为偶数时,

$$a_k = c_k;$$

当 k 为奇数时,

$$\begin{aligned} a_k &= \sqrt{\frac{2}{\pi}}\sigma^k \int_0^{\infty} x^k e^{-\frac{x^2}{2}}\, dx = \sqrt{\frac{2}{\pi}}\sigma^k 2^{\frac{k-1}{2}}\Gamma\Big(\frac{k+1}{2}\Big) \\ &= \sqrt{\frac{2}{\pi}} 2^{\frac{k-1}{2}}\Big(\frac{k-1}{2}\Big)\sigma^k. \end{aligned}$$

例 4 数学期望可能不存在. 设 ξ 的密度为 [1]

$$f(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}. \tag{29}$$

由于

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{|x|}{1+x^2}\, dx = \infty,$$

图 1

故 $E\xi$ 不存在，从而高阶矩及方差也不存在.

(29) 中的密度可如下一般化：取二常数 μ 及 $\lambda > 0$, 称密度为

$$f_{\lambda,\mu}(x) = \frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2 + (x-\mu)^2} \tag{30}$$

的分布为柯西 (Cauchy) 分布.

设直线 AC 可绕定点 A 旋转， AB 垂直 $x-$ 轴，长为 λ, B 的横坐标为 μ; 又设角度 $\theta = \angle BAC$ 是随机变数，在 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 中均匀分布. 则 C 点的横坐标 ξ 也是随机变数 (见图 1). 显然

$$\xi = \mu + \lambda \operatorname{tg}\theta,$$

由 §2.8 定理 1, 易见 ξ 的密度由 (30) 给出.

例 5 设 $\widetilde{E} = E^{\infty}$ 为可列重贝努里试验，事件 A 在每次简单试验 E 中出现的概率为 $p, 0 < p < 1$. 定义

$$\xi_k = \begin{cases} 1, & \text{如第 } k \text{ 次试验时 } A \text{ 出现} \\ 0, & \text{反之} \end{cases} \tag{31}$$

则 $\sum\limits_{k=1}^{n}\xi_k$ 是前 n 次试验中 A 出现的总数，我们记得 $\sum\limits_{k=1}^{n}\xi_k$ 有二项分布 $B(n,p)$, 即

$$P\Big(\sum_{k=1}^{n}\xi_k = j\Big) = \binom{n}{j} p^j q^{n-j}, \quad (j = 1, 2, \cdots, n; q = 1-p) \tag{32}$$

以后称由 (31) 定义的 $\{\xi_k\}$ 为连系于贝努里试验 $\widetilde{E}$ 的随机变数列；注意 $\{\xi_k\}$ 是独立的.

[1] 设 X_1 与 X_2 为独立，有相同 $N(0,\sigma)$ 分布，则 X_1/X_2 有此密度，参看 90 页. 关于柯西分布的另一些性质，见 194 页 25 题.

现在设 N 是一具有参数为 $\lambda > 0$ 的普阿松分布的随机变数，而且 $\{N, \xi_1, \xi_2, \cdots\}$ 是相互独立的，ξ_k 由 (31) 定义. 我们来求复合随机变数

$$\eta = \sum_{k=1}^{N} \xi_k \tag{33}$$

的分布与数学期望. 注意，(33) 中的 N 是随机变数而不是普通的常数，这样的和称为随机和. 显然，η 只能取非负整数值 (如 $N=0$, 令 $\eta = 0$).

由全概率公式

$$P(\eta = j) = \sum_{n=0}^{\infty} P(\eta = j|\ N = n) P(N = n),\ (j = 0, 1, 2, \cdots), \tag{34}$$

然而 $P(N=n) = e^{-\lambda}\frac{\lambda^n}{n!}$; 又由独立性得

$$\begin{aligned} P(\eta = j|N = n) &= \frac{P(\eta = j, N = n)}{P(N = n)} = \frac{P\left(\sum\limits_{k=1}^{n} \xi_k = j, N = n\right)}{P(N = n)} \\ &= \frac{P\left(\sum\limits_{k=1}^{n} \xi_k = j,\right) P(N = n)}{P(N = n)} = \binom{n}{j} p^j q^{n-j}. \end{aligned} \tag{35}$$

代入 (34), 并注意 $\binom{n}{j} = 0\ (j > n)$, 即得

$$\begin{aligned} P(\eta = j) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \cdot \binom{n}{j} p^j (1-p)^{n-j} \\ &= \frac{e^{-\lambda} p^j}{j!} \sum_{n=j}^{\infty} \frac{\lambda^n (1-p)^{n-j}}{(n-j)!} = \frac{e^{-\lambda} p^j \lambda^j}{j!} \sum_{n=j}^{\infty} \frac{[\lambda(1-p)]^{n-j}}{(n-j)!} \\ &= \frac{e^{-\lambda} p^j \lambda^j}{j!} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^j}{j!} \end{aligned} \tag{36}$$

故得知 η 是具有参数为 λp 的普阿松分布. 由例 2, 知 $E\eta = \lambda p$, $D\eta = \lambda p$. 注意 $E\xi_i = p$, $EN = \lambda$, 故 $E\eta = EN \cdot E\xi_j$; 这式具有明显的直观意义.

例 6 仍考虑例 5 中的 $\widetilde{E} = E^{\infty}$, 试求第 n 次出现 A 时的试验次数的数学期望.

解 以 ζ_1 表第一次出现 A 的试验次数，以 $\zeta_k (k > 1)$ 表第 $k-1$ 次出现 A 的下一次试验起到第 k 次出现 A 时止所做的试验次数 (包括第 k 次出现 A 的那一次), 显然，第 n 次出现 A 时的试验次数 ζ 为

$$\zeta = \sum_{k=1}^{n} \zeta_k.$$

$\zeta_k (k = 1, 2, \cdots)$ 有相同的几何分布 (见 §2.3 例 4), 故

$$\begin{aligned} E\zeta &= \sum_{k=1}^{n} E\zeta_k = nE\zeta_1 \\ &= n \cdot \sum_{k=1}^{\infty} kq^{k-1} p = n \cdot \frac{p}{(1-q)^2} = \frac{n}{p}. \end{aligned} \tag{37}$$

由例 6 可见，如射击独立进行，每次射中的概率为 p, $(p>0)$, 则到第 n 次射中时，平均射出的子弹数为 $\frac{n}{p}$. 例如，若 $p=\frac{1}{2}$, 则 $E\zeta=2n$, 这与直观符合.

***例 7 (广义蒲丰问题)** 平面上画有等距离为 a 的一些平行线，向平面任意投一凸形铁丝圈，其长为 s. 试求此圈与平行线的交点个数的数学期望.

解 将铁丝圈分成若干小段，其中一段长为 $l_k(<a)$. 当 l_k 充分小时，可把这小段看成一直线段. 根据 §1.1 例 8, 可见这小段与某平行线相交的概率为 $2l_k/\pi a$. 定义

$$\xi_k=\begin{cases}1, & \text{如第 } k \text{ 小段与某平行线相交,}\\ 0, & \text{如第 } k \text{ 小段与任一平行线都不相交.}\end{cases}$$

显然，$\sum\limits_k \xi_k$ 等于铁丝圈与平行线的交点数. 于是

$$\begin{aligned}E\Big(\sum_k\xi_k\Big)&=\sum_k E\xi_k=\sum_k\left[1\cdot\frac{2l_k}{\pi a}+0\cdot\left(1-\frac{2l_k}{\pi a}\right)\right]\\&=\sum_k\frac{2l_k}{\pi a}=\frac{2s}{\pi a}.\end{aligned}$$

利用这个结果，还可解决下列问题：设上述铁丝圈是闭的凸形圈，而且充分小. 试求它与某平行线相交的概率 p. 由于这个圈的交点数不是 2 就是 0(相切时认为交点有两个), 故交点数的数学期望为 $2p+0\cdot(1-p)=2p$. 但上面又算出它应等于 $\frac{2s}{\pi a}$. 比较这两个结果即得

$$p=\frac{s}{\pi a}.$$

注意此式与 s 的形状无关，现在取铁丝圈为一椭圆，而且它越来越扁，最后趋向一长度为 $l=\frac{s}{2}$ 的针. 于是上式化为

$$p=2\cdot\frac{s}{2}/\pi a=2l/\pi a.$$

这正是 §1.1(8) 式.

*(六) **关于数学期望的定义的补充说明**. 设 ξ 的分布函数为 $F_\xi(x)$, 又 $\eta=g(\xi)$, 的分布函数为 $F_\eta(x)$, 这里 $g(x)(x\in R_1)$ 为波雷耳可测函数. 如果 η 的数学期望存在，那么 $E\eta=\int_{-\infty}^{\infty}x\,dF_\eta(x)$; 另方面， $Eg(\xi)=\int_{-\infty}^{\infty}g(x)\,dF_\xi(x)$. 既然 $\eta=g(\xi)$, 自然应该有

$$\int_{-\infty}^{\infty}x\,dF_\eta(x)=\int_{-\infty}^{\infty}g(x)\,dF_\xi(x). \tag{38}$$

否则，数学期望的定义中就含有内在的矛盾. 的确，可以证明：

定理 4 (38) 中如有一积分存在，那么另一积分也存在，而且二者相等.

证 由于 x 连续，故

$$\int_{-\infty}^{\infty}x\,dF_\eta(x)=\lim_{n\to\infty}\sum_i x_i^{(n)}[F_\eta(x_i^{(n)}-F_\eta(x_{i-1}^{(n)})], \tag{39}$$

其中 $\cdots<x_{-1}^{(n)}<x_0^{(n)}<x_1^{(n)}<\cdots$ 是任一列分点， $\lim\limits_{i\to\infty}x_i^{(n)}=\infty$, $\lim\limits_{i\to-\infty}x_i^{(n)}=-\infty$, 而且 $\lim\limits_{n\to\infty}\max\limits_i(x_i^{(n)}-x_{i-1}^{(n)})=0$. 注意， (39) 中积分存在的条件是右方级数对任意如上的分点列 $\{x_i^{(n)}\}$ $(n=1,2,\cdots)$ 绝对收敛.

另方面，根据勒贝格司梯阶积分的性质，有

$$\int_{-\infty}^{\infty} g(x)\,dF_\xi(x) = \lim_{n\to\infty}\sum_i x_i^{(n)} F_\xi\{x:\ x_{i-1}^{(n)} < g(x) \leqslant x_i^{(n)}\}. \tag{40}$$

(40) 中积分存在的充要条件是右方级数对任意如上的 $\{x_i^{(n)}\}$ 绝对收敛. 然而

$$\begin{aligned}&F_\eta(x_i^{(n)}) - F_\eta(x_{i-1}^{(n)}) = P(x_{i-1}^{(n)} < \eta \leqslant x_i^{(n)}) = P(x_{i-1}^{(n)} < g(\xi) \leqslant x_i^{(n)})\\ &= P(\xi \in \{x:\ x_{i-1}^{(n)} < g(x) \leqslant x_i^{(n)}\}) = F_\xi(x:\ x_{i-1}^{(n)} < g(x) \leqslant x_i^{(n)}).\end{aligned}$$

由此及 (39)(40) 即得证 (38). □

定理 4 可推广到 n 维情形. 设 $\xi = (\xi_1,\cdots,\xi_n)$ 的分布函数为 $F_\xi(x_1,\cdots,x_n)$, $\eta = g(\xi_1,\cdots,\xi_n)$ 的分布函数为 $F_\eta(y)$, $(y \in R_1)$, 这里 $g(x_1,\cdots,x_n)$ 为 n 元波雷耳可测函数. 我们有

定理 5

$$\int_{-\infty}^{\infty} y\,dF_\eta(y) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\cdots,x_n)\,dF_\xi(x_1,\cdots,x_n) \tag{41}$$

它的证明完全与定理 4 一样，只要把那里的 x 理解为 R_n 中的点 $x = (x_1,\cdots,x_n)$, 把 ξ 理解为 n 维随机向量 $\xi = (\xi_1,\cdots,\xi_n)$ 就行了.

***(七) 数学期望的几何性质.** 我们来证明两个几何性质，这对计算 $E\xi$ 有时很方便.

设 ξ 的分布函数为 $F(x)$, 数学期望 $E\xi$ 存在，即

$$\int_{-\infty}^{\infty} |x|\,dF(x) < \infty, \tag{42}$$

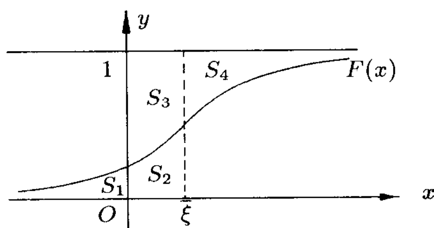


图 2

简记 $\bar{\xi} = E\xi$. 则有

性质 A

$$\int_{-\infty}^{\bar{\xi}} F(x)\,dx = \int_{\bar{\xi}}^{\infty} [1 - F(x)]\,dx,$$

(亦即图 2 中诸面积间有等式 $s_1 + s_2 = s_4$)

性质 B

$$\bar{\xi} = \int_0^{\infty} [1 - F(x)]\,dx - \int_{-\infty}^{0} F(x)\,dx,$$

(亦即 $\bar{\xi} = s_3 + s_4 - s_1$)

证 A

$$\begin{aligned}\int_{-\infty}^{\bar{\xi}} F(x)\,dx &= xF(x)\Big|_{-\infty}^{\bar{\xi}} - \int_{-\infty}^{\bar{\xi}} xdF(x)\\ &= \bar{\xi}F(\bar{\xi}) - \lim_{x\to-\infty} xF(x) - \int_{-\infty}^{\bar{\xi}} xdF(x),\end{aligned}$$

由 (42)

$$\lim_{x\to-\infty}|xF(x)| = \lim_{x\to-\infty}\left|x\int_{-\infty}^{x} dF(y)\right|$$
$$\leqslant \lim_{x\to-\infty}\left|\int_{-\infty}^{x} ydF(y)\right| \leqslant \lim_{x\to-\infty}\int_{-\infty}^{x}|y|dF(y)=0,$$

故

$$\int_{-\infty}^{\bar{\xi}} F(x)\,dx = \bar{\xi}F(\bar{\xi}) - \int_{-\infty}^{\bar{\xi}} xF(x)\,dx. \tag{43}$$

类似得 A 的右式为

$$\int_{\bar{\xi}}^{\infty}[1-F(x)]\,dx = \lim_{x\to\infty}x[1-F(x)] - \bar{\xi}[1-F(\bar{\xi})] + \int_{\bar{\xi}}^{\infty} xdF(x).$$

但由 (42)

$$0 \leqslant \lim_{x\to\infty}x[1-F(x)] = \lim_{x\to\infty}x\int_{x}^{\infty}dF(y) < \lim_{x\to\infty}\int_{x}^{\infty}ydF(y)=0,$$

故

$$\int_{\bar{\xi}}^{\infty}[1-F(x)]\,dx = \left(-\bar{\xi}+\int_{\bar{\xi}}^{\infty}xdF(x)\right)+\bar{\xi}F(\bar{\xi})$$
$$= -\int_{-\infty}^{\bar{\xi}}xdF(x)+\bar{\xi}F(\bar{\xi}).$$

由此及 (43) 即得证 A.

证 B 记

$$s_1 = \int_{-\infty}^{0}F(x)\,dx,\quad s_2=\int_{0}^{\bar{\xi}}F(x)\,dx,$$
$$s_3 = \int_{0}^{\bar{\xi}}[1-F(x)]\,dx,\quad s_4=\int_{\bar{\xi}}^{\infty}[1-F(x)]\,dx.$$

(见图 2) 由 A 得 $s_1+s_2=s_4$. 故

$$\int_{0}^{\infty}[1-F(x)]\,dx - \int_{-\infty}^{0}F(x)\,dx$$
$$= s_3+s_4-s_1 = s_3+(s_1+s_2)-s_1 = s_2+s_3 = \bar{\xi}\cdot 1 = \bar{\xi}.$$

这里假定了 $\bar{\xi}\geqslant 0$. 对 $\bar{\xi}<0$ 的情况可类似证明.

§2.10 随机向量的数字特征

(一) 预备知识. 上节中对随机变数引进的一些数字特征可以推广到随机向量. 先作些准备工作.

先考虑定义在概率空间 $(\Omega,\mathcal{F},P)$ 上的二随机变数 $\xi_1(\omega),\xi_2(\omega)$. 说它们相等是指：对一切 $\omega\in\Omega$,

$$\xi_1(\omega)=\xi_2(\omega). \tag{1}$$

然而在概率论地绝大多数问题中，只需 "$\xi_1(\omega),\xi_2(\omega)$ 几乎处处相等" 就够了，这句话的精确意思是说：

$$P(\omega:\xi_1(\omega)=\xi_2(\omega))=1; \tag{2}$$

关系式 (2) 通常记为

$$\xi_1=\xi_2\quad(\text{a.s.}). \tag{3}$$

现在证明：如 (2) 成立，则 ξ_1 与 ξ_2 有相同的分布函数. 因此，任何一个由分布函数确定的性质或由分布函数确定的数字特征，对 ξ_1 与 ξ_2 都是相同的；例如，由 §2.9(4) 可见，如果 ξ_1 的数学期望存在，那么 ξ_2 的也存在而且二者相同，对方差及其它各阶矩也如此. 由 (2) 得 $P(\xi_1\neq\xi_2)=0$, 故

$$\begin{aligned}P(\xi_1\leqslant x)&=P([\xi_1\leqslant x]\cap[\xi_1\neq\xi_2])+P([\xi_1\leqslant x]\cap[\xi_1=\xi_2])\\&=P([\xi_1\leqslant x]\cap[\xi_1=\xi_2])\\&=P([\xi_2\leqslant x]\cap[\xi_1=\xi_2])=P(\xi_2\leqslant x).\end{aligned} \tag{4}$$

这得证 ξ_1,ξ_2 有相同的分布函数.

一般地，设有某性质 C, 它涉及基本事件 ω, 我们说，C 几乎处处成立. 或者说，C 以概率 1(记为 a.s.) 成立，是指使 C 成立的 ω 所构成的集 A_C 的概率为 1, 即 $P(A_C)=1$.

例如，在 (2) 中，C 表性质 "$\xi_1(\omega)=\xi_2(\omega)$",$A_C=(\omega:\xi_1(\omega)=\xi_2(\omega))$; 又例如

$$\xi\geqslant 0\ (\text{a.s.})\ \text{表示} P(\omega:\xi(\omega)\geqslant 0)=1;$$
$$\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\ (\text{a.s.})\ \text{表示} P(\omega:\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega))=1.$$

引理 1 (马尔科夫 [Марков] 不等式) 设随机变数 ξ 有 r 阶绝对矩，$E|\xi|^r<\infty,(r>0)$, 则对任意 $\varepsilon>0$ 有

$$P(|\xi|\geqslant\varepsilon)\leqslant\frac{E|\xi|^r}{\varepsilon^r}. \tag{5}$$

证 设 ξ 的分布为 $F(A)$, $(A\in\mathcal{B}_1)$, 分布函数为 $F(x)(x\in R_1)$. 则

$$\begin{aligned}P(|\xi|\geqslant\varepsilon)&=F(|x|\geqslant\varepsilon)=\int_{|x|\geqslant\varepsilon}dF(x)\\&\leqslant\int_{|x|\geqslant\varepsilon}\frac{|x|^r}{\varepsilon^r}\,dF(x)\leqslant\frac{1}{\varepsilon^r}\int_{-\infty}^{\infty}|x|^r\,dF(x)=\frac{E|\,\xi|^r}{\varepsilon^r}.\quad\square\end{aligned}$$

(5) 式的特殊情形：取 $r=2$, 并以 $\xi-E\xi$ 代替 ξ, 得

$$P(|\xi-E\xi|\geqslant\varepsilon)\leqslant\frac{D\xi}{\varepsilon^2},$$

此式称为车贝晓夫 (Чебышев) 不等式.

引理 2 对随机变数 ξ, $\xi=c$ (a.s.) 的充要条件是 $D\xi=0$ (c 为常数).

证 必要性已在 §2.9 定理 2(i) 中证明.

反之，设 $D\xi=0$, 以 $\xi-E\xi$ 代替 (5) 中 ξ, 并令 $r=2$, 则对任意 $\varepsilon>0$

$$P(|\xi-E\xi|\geqslant\varepsilon)\leqslant\frac{D\xi}{\varepsilon^2}=0.$$

由于事件 $\left(|\xi-E\xi|\geqslant\frac{1}{n}\right)\uparrow(|\xi-E\xi|\neq 0)$, 故

$$P(|\xi-E\xi|\neq 0)=\lim_{n\to\infty}P\left(|\xi-E\xi|\geqslant\frac{1}{n}\right)=0,$$
$$P(\xi=E\xi)=1-P(|\xi-E\xi|\neq 0)=1.\quad\square$$

(二) 相关系数. 设已给 n 维随机向量

$$\xi(\omega)=(\xi_1(\omega),\cdots,\xi_n(\omega)).$$

如果 $E\xi_i(i=1,\cdots,n)$ 都存在，称 n 维 (常数) 向量 $(E\xi_1,\cdots,E\xi_n)$ 为 ξ 的数学期望，并记为

$$E\xi=(E\xi_1,\cdots,E\xi_n). \tag{6}$$

如果

$$b_{jk}=E(\xi_j-E\xi_j)(\xi_k-E\xi_k) \tag{7}$$

存在，则称 b_{jk} 为 ξ_j 与 ξ_k 的二阶混合中心矩，而 n 级矩阵

$$B=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n}\\ b_{21} & b_{22} & \cdots & b_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ b_{n1} & b_{n2} & \cdots & b_{nn}\end{pmatrix} \tag{8}$$

则称为 ξ 的协方差矩阵，它的行列式记为 $|B|$, 注意

$$b_{jj}=D\xi_j.$$

矩阵 B 具有下列性质：

1° 对称性： $b_{jk}=b_{kj}$ (一切 $j,k=1,\cdots,n$).

2° 非负定性：对任意实数 $\eta_1,\cdots,\eta_n$, 有

$$\sum_{j,k=1}^{n}b_{jk}\eta_j\eta_k\geqslant 0 \tag{9}$$

(比较 §2.6 例 7). 实际上， 1° 由 (7) 而明显；又

$$\sum_{j,k=1}^{n}b_{jk}\eta_j\eta_k=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\sum_{j=1}^{n}\eta_j(x_j-E\xi_j)\right\}^2dF(x_1,\cdots,x_n)\geqslant 0.$$

由 (9) 及二次型的理论知

$$|B|\geqslant 0. \tag{10}$$

现在对任意两个随机变数 ξ_1,ξ_2 来引进一个重要的数字特征. 设 $b_{11}=D\xi_1>0, b_{22}=D\xi_2>0$, 称

$$r_{12}=\frac{b_{12}}{\sqrt{b_{11}b_{22}}}=\frac{E(\xi_1-E\xi_1)(\xi_2-E\xi_2)}{\sqrt{E(\xi_1-E\xi_1)^2\cdot E(\xi_2-E\xi_2)^2}} \tag{11}$$

为 ξ_1 与 ξ_2 的相关系数.

以 $F_{12}(x_1,x_2)$ 表 (ξ_1,ξ_2) 的分布函数，由布尼亚科夫斯基不等式得

$$\begin{aligned}|b_{12}|&=\left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-E\xi_1)(x_2-E\xi_2)\,dF_{12}(x_1,x_2)\right|\\&\leqslant\sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-E\xi_1)^2\,dF_{12}(x_1,x_2)}\sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_2-E\xi_2)^2\,dF_{12}(x_1,x_2)}\\&=\sqrt{b_{11}b_{22}},\end{aligned} \tag{12}$$

可见当 ξ_1,ξ_2 的方差有穷且不等于 0 时， r_{12} 有定义而且有穷.

相关系数且有下列性质 1° − 4°.

1° $-1\leqslant r_{12}\leqslant 1$. 这由 (11)(12) 直接推出.

2° 如 ξ_1,ξ_2 独立，则 $r_{12}=0$. 事实上，由 §2.9 定理 1(iii)

$$E(\xi_1-E\xi_1)(\xi_2-E\xi_2)=E\xi_1E\xi_2-E\xi_1E\xi_2-E\xi_1E\xi_2+E\xi_1E\xi_2=0.$$

3° $|r_{12}|=1$ 的充要条件是： ξ_1 与 ξ_2 以概率 1 线性相关；换句话说，存在常数 a,b, 使

$$\xi_1=a\xi_2+b\ (a.s.). \tag{14}$$

证 充分：由 (14) 得

$$\begin{aligned}b_{12}&=E(\xi_1-E\xi_1)(\xi_2-E\xi_2)=E(a\xi_2+b-aE\xi_2-b)(\xi_2-E\xi_2)\\&=aE(\xi_2-E\xi_2)^2=ab_{22},\\b_{11}&=E(\xi_1-E\xi_1)^2=E(a\xi_2-aE\xi_2)^2=a^2b_{22},\\r_{12}&=\frac{b_{12}}{\sqrt{b_{11}b_{22}}}=\frac{ab_{22}}{|a|b_{22}}.\end{aligned}$$

由于已假定 $b_{11}>0, b_{22}>0$, 故 $a\neq 0$, 于是 $|r_{12}|=1$. 还可看出如 $a>0$, 则 $r_{12}=1$, 如 $a<0$, 则 $r_{12}=-1$.

必要：易见

$$D\left(\frac{\xi_1}{\sqrt{b_{11}}}\pm\frac{\xi_2}{\sqrt{b_{22}}}\right)=2(1\pm r_{12})$$

由引理 2 知：存在常数 c, 使

$$\left.\begin{aligned}&\text{当}r_{12}=1,\ \frac{\xi_1}{\sqrt{b_{11}}}-\frac{\xi_2}{\sqrt{b_{22}}}=c\ (\text{a.s.}),\\&\text{当}r_{12}=-1,\ \frac{\xi_1}{\sqrt{b_{11}}}+\frac{\xi_2}{\sqrt{b_{22}}}=c\ (\text{a.s.}).\end{aligned}\right\}.\ \square \tag{15}$$

由性质 2°, 3° 可见：当 ξ_1, ξ_2 独立时 $|r_{12}|$ 达到极小值 0; ξ_1, ξ_2 线性相关时它达到极大值 1. 而且后一结论之逆也正确. 这说明 r_{12} 在一定程度上表达了 ξ_1, ξ_2 间的相关程度，所以叫它为相关系数.

如果 $r_{12}=0$, 就称 ξ_1, ξ_2 不相关.

由 (15) 看出，如 $r_{12}=1$,(14) 中的 $a>0$, ξ_2 值变大时， ξ_1 也变大；如 $r_{12}=-1$, 同理知 ξ_2 变大时， ξ_1 变小，前一种相关称为正相关而后一种称为负相关.

注意 2° 之逆不正确：由不相关一般推不出独立性，因而 "不相关" 与 "独立" 是两不同的概念.

例 1 设 ξ_1, ξ_2 的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & \text{如}x^2+y^2\leqslant 1,\\ 0, & \text{如}x^2+y^2>1.\end{cases} \tag{16}$$

由 f 的对称性 (或由直接计算) 可见 ξ_1, ξ_2 的二阶混合中心矩为 0, 故其相关系数也为 0; 但 ξ_1, ξ_2 不独立，例如

$$P\Big(\xi_1>\frac{1}{\sqrt{2}}, \xi_2>\frac{1}{\sqrt{2}}\Big)\neq P\Big(\xi_1>\frac{1}{\sqrt{2}}\Big)P\Big(\xi_2>\frac{1}{\sqrt{2}}\Big).$$

因为显然右方大于 0, 而左方由于 $\xi_1^2+\xi_2^2>1$ 而等于 0.

虽然如此，对二维正态分布， 2° 之逆却是正确的.

例 2 如果 ξ_1, ξ_2 有二维正态分布，密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]} \tag{17}$$

则

$$E\xi_1=a, E\xi_2=b, D\xi_1=\sigma_1^2, D\xi_2=\sigma_2^2, \tag{18}$$

$$r_{12}=r. \tag{19}$$

而且 ξ_1, ξ_2 的独立性等价于不相关性.

实际上，由 §2.6 例 6 知 ξ_1, ξ_2 的密度函数分别为

$$f_1(x)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma_1^2}},\ f_2(y)=\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(y-b)^2}{2\sigma_2^2}}. \tag{20}$$

由 §2.9 例 3 知 (18) 成立，又

$$\begin{aligned}b_{12}=b_{21}&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a)(y-b)f(x,y)\,dxdy\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{\infty}e^{-\frac{(y-b)^2}{2\sigma_2^2}}\,dy\cdot\int_{-\infty}^{\infty}(x-a)(y-b)\\&\quad\cdot\exp\Big\{-\frac{1}{2(1-r^2)}\Big(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\Big)^2\Big\}\,dx.\end{aligned}$$

作变换 $z=\frac{1}{\sqrt{1-r^2}}\left(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\right)$, $\omega=\frac{y-b}{\sigma_2}$, 上式化为

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-r^2}\omega z+r\sigma_1\sigma_2\omega^2)e^{-\frac{\omega^2}{2}-\frac{z^2}{2}}dzd\omega$$

$$=\frac{r\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}\omega^2e^{-\frac{\omega^2}{2}}d\omega\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz+\frac{\sigma_1\sigma_2\sqrt{1-r^2}}{2\pi}\int_{-\infty}^{\infty}\omega e^{-\frac{\omega^2}{2}}d\omega\int_{-\infty}^{\infty}ze^{-\frac{z^2}{2}}dz,$$

其中第一，第二积分都等于 $\sqrt{2\pi}$, 第三，第四积分都等于 0, 故

$$b_{12}=r\sigma_1\sigma_2.$$

于是

$$r_{12}=\frac{b_{12}}{\sqrt{D\xi_1D\xi_2}}=r.$$

由 2° 及 §2.7 例 1 立得： ξ_1,ξ_2 独立的充要条件是它们的相关系数 $r_{12}=0$. 换句话说：

4° 对于有二维正态分布的随机变数 ξ_1,ξ_2, 独立性与不相关性是等价的.

(三) **中数, 众数**. 至今看到的数字特征都是 §2.9(5) 式 (或 (9) 式) 的特殊情形, 它们在下章中将起重要作用. 数学期望与方差服从简单的加法运算与乘法运算法则 (见 §2.9(10),(14) 与 (11)), 这是它们的优点；缺点是它们并不是对一切随机变数都有定义，而且即使有定义也未必能轻易地求出其值.

现在来引进一些新的数字特征，后者能避免这些缺点；然而，可惜一般又没有上述的优点.

对任意随机变数 ξ, 满足下二式

$$P(\xi\leqslant x)\geqslant\frac{1}{2},\quad P(\xi\geqslant x)\geqslant\frac{1}{2}\tag{21}$$

的 x 称为 ξ 的中数，记为 $x_{\frac{1}{2}}$, 它是反映集中位置的一个数字特征. 中数总存在，但可以不唯一.

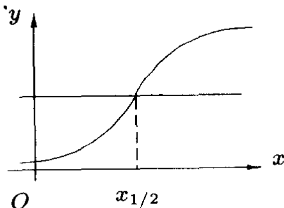


图 1

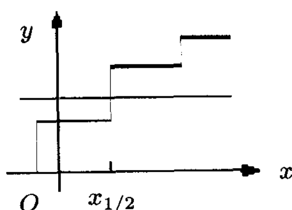


图 2

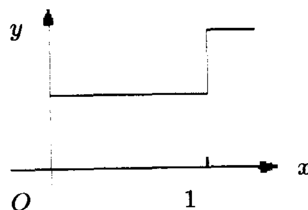


图 3

画出 ξ 的分布函数 $F(x)$ 的图. 如果 $F(x)$ 连续，那么 $x_{\frac{1}{2}}$ 是方程

$$F(x)=\frac{1}{2}\tag{22}$$

的解 (图 1). 如果它有跳跃点 (见图 2), 用平行于 y 轴的直线联接后，得一连续曲线，它与直线 $y=\frac{1}{2}$ 的交点横坐标即 $x_{\frac{1}{2}}$, 由于交点可以不唯一，故可有许多个 $x_{\frac{1}{2}}$; 例如图 3 表密度为 $\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$ 的分布函数， $x_{\frac{1}{2}}$ 充满了区间 $[0,1]$.

设 $F(x)$ 有密度 $f(x)$, 而且 $f(x)$ 的图形关于直线 $x=a$ 对称，即 $f(a+x)=f(a-x)$, 那么 $x_{\frac{1}{2}}=a$; 如果这时数学期望 m 存在，则 $m=a=x_{\frac{1}{2}}$. 特别，对于正态分布 $N(a,\sigma)$, $m=a=x_{\frac{1}{2}}$.

一般地，对于任意常数 $r, 0<r<1$, 满足

$$P(\xi \leqslant x) \geqslant r, \quad P(\xi \geqslant x) \geqslant 1-r \tag{23}$$

的 x 称为 $r-$ 位数，记为 x_r, 对 x_r 可进行类似的讨论. 称 $x_{\frac{3}{4}}-x_{\frac{1}{4}}$ 为内极差. 如 ξ 的分布 $F(A),(A\in\mathcal{B}_1)$ 是连续型的，显然，$F([x_{\frac{1}{4}},x_{\frac{3}{4}}])=\frac{1}{2}$; 因此，如 $x_{\frac{3}{4}}-x_{\frac{1}{4}}$ 愈小，分布的半质量也越集中，故 $x_{\frac{3}{4}}-x_{\frac{1}{4}}$ 是表示集中程度的数字特征. 类似可定义一般的 $x_{r_2}-x_{r_1}, r_2>r_1$.

对于一些特殊的随机变数 ξ, 还可定义另一反映集中位置的数字特征：众数. 如 ξ 是离散型的，称 ξ 的最可能值 (见 §2.3(17) 为众数；如 ξ 有连续的密度函数 $f(x)$, 称使 $f(x)$ 达到最大值的点 x 为众数，众数也可不唯一.

对 $N(a,\sigma)$, 显然众数，数学期望，中数都重合为 a.

例 3 对具有密度为 $f(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$ 的柯西分布，有

$$F(x)=\frac{1}{\pi}\int_{-\infty}^{x}\frac{d\mu}{1+\mu^2}=\frac{1}{2}+\frac{1}{\pi}\operatorname{arctg}x.$$

由于 $F(-1)=\frac{1}{4},F(0)=\frac{1}{2},F(1)=\frac{3}{4}$, 故中数为 0, 而 $x_{\frac{3}{4}}-x_{\frac{1}{4}}=1-(-1)=2$, 又因 $f(x)$ 关于原点对称，连续，在 $x=0$ 取最大值 $\frac{1}{\pi}$, 故由此也可看出 $x_{\frac{1}{2}}=0$, 而且众数重合于 $x_{\frac{1}{2}}$.

§2.11 特 征 函 数

(一) 特征函数的定义和性质.

除了一些特殊的分布 (如二项分布，普阿松分布，正态分布等) 被它的数学期望和方差所唯一决定外，在一般情况下，数学期望，方差等只能粗略地反映分布函数的某些性质. 能够完全刻画分布函数的是它的特征函数，后者是本节研究的对象. 特征函数有时比分布函数更便于应用，例如，矩的计算对分布函数是积分而特征函数则是微分；研究独立随机变数和的分布时，用分布函数是求卷积，而用特征函数则化为简单的乘法 (见下面性质 (iv) (v)) 在极限定理的研究中，特征函数尤其起着重要的工具作用.

设已给分布函数 $F(x),x\in R_1$, 它的特征函数 $f(t)(t\in R_1)$ 定义为 [1]

$$f(t)=\int_{-\infty}^{\infty}e^{itx}dF(x) \tag{1}$$

由于 $|e^{itx}|=1$ 而且 F 有界，故 $f(t)$ 对一切 $t\in R_1$ 有定义.

如 F 是离散型的，密度为 $\binom{a_0\ a_1\ \cdots}{p_0\ p_1\ \cdots}$, 则

$$f(t)=\sum_j p_j e^{ita}.$$

[1] $f(t)=\int_{-\infty}^{\infty}\cos tx\,dF(x)+i\int_{-\infty}^{\infty}\sin tx\,dF(x)$.

如 F 是连续型的，密度为 $g(x)$, 则

$$f(t)=\int_{-\infty}^{\infty}e^{itx}g(x)\,dx.$$

随机变数 ξ 的特征函数 $f(t)$ 定义为它的分布函数 $F(x)$ 的特征函数，即

$$f(t)=\int_{-\infty}^{\infty}e^{itx}dF(x)=Ee^{it\xi}. \tag{2}$$

试述特征函数 $f(t)$ 的性质：

(i) $f(t)$ 在 R_1 上均匀连续，而且

$$|f(t)|\leqslant f(0)=1. \tag{3}$$

$$f(-t)=\overline{f(t)}\ \text{(横线表共轭复数)}. \tag{4}$$

证

$$\begin{aligned}|f(t+h)-f(t)|&=\left|\int_{-\infty}^{\infty}e^{itx}(e^{ihx}-1)\,dF(x)\right|\leqslant\int_{-\infty}^{\infty}|e^{ihx}-1|\,dF(x)\\&\leqslant\int_{-a}^{a}|e^{ihx}-1|dF(x)+2\int_{|x|\geqslant a}dF(x).\end{aligned} \tag{5}$$

对任给 $\varepsilon>0$, 取 a 充分大，使

$$\int_{|x|\geqslant a}dF(x)<\frac{\varepsilon}{4}; \tag{6}$$

再取 $\delta>0$, 使 $|h|<\delta$, 对一切 $x\in[-a,a]$ 均匀地有

$$|e^{ihx}-1|=\left|e^{i\frac{hx}{2}}\left(e^{i\frac{hx}{2}}-e^{-i\frac{hx}{2}}\right)\right|=2\left|\sin\frac{hx}{2}\right|<\frac{\varepsilon}{2},$$

由此及 (5)(6) 即得 $|f(t+h)-f(t)|<\varepsilon$, 只要 $|h|<\delta$, 此得证 $f(t)$ 的均匀连续性.

其次，

$$|f(t)|=\left|\int_{-\infty}^{\infty}e^{itx}\,dF(x)\right|\leqslant\int_{-\infty}^{\infty}|e^{itx}|\,dF(x)=\int_{-\infty}^{\infty}dF(x)=1=f(0).$$

$$f(-t)=\int_{-\infty}^{\infty}e^{-itx}\,dF(x)=\overline{\int_{-\infty}^{\infty}e^{itx}\,dF(x)}=\overline{f(t)}\ \ \square$$

(ii) $f(t)$ 是非负定的：对任意正整数 n, 任意 $t_i\in R_1$ 及复数 λ_i, $i=1,\cdots,n$, 有

$$\sum_{i,j=1}^{n}f(t_i-t_j)\lambda_i\overline{\lambda}_j\geqslant 0. \tag{7}$$

证

$$\begin{aligned}\sum_{i,j=1}^{n}f(t_i-t_j)\lambda_i\overline{\lambda}_j&=\int_{-\infty}^{\infty}\sum_{i,j=1}^{n}(e^{it_ix}e^{-it_jx}\lambda_i\overline{\lambda}_j)\,dF(x)\\&=\int_{-\infty}^{\infty}\left|\sum_{i=1}^{n}\lambda_ie^{it_ix}\right|^2dF(x)\geqslant 0\ \square\end{aligned}$$

设随机变数的分布函数为 $F(x)$, 特征函数为 $f(t)$, 有必要强调 ξ 时，记为 $F_\xi(x)$, $f_\xi(t)$.

(iii) 设 $\eta = \alpha\xi + b$, a, b 为二常数，则

$$f_\eta(t) = e^{ibt} f_\xi(at). \tag{8}$$

证 $f_\eta(t) = Ee^{it\eta} = Ee^{it(a\xi+b)} = e^{ibt}Ee^{iat\xi} = e^{ibt}f_\xi(at)$ □

(iv) 设 η 是 n 个独立随机变数 $\xi_1, \cdots, \xi_n$ 的和

$$\eta = \sum_{i=1}^{n} \xi_i. \tag{9}$$

则 [1]

$$f_\eta(t) = \prod_{i=1}^{n} f_{\xi_i}(t). \tag{10}$$

证 设 $n = 2$, 有

$$\begin{aligned} f_\eta(t) &= Ee^{it(\xi_1+\xi_2)} = Ee^{it\xi_1}e^{it\xi_2} \\ &= E(\cos t\xi_1 + i\sin t\xi_1)(\cos t\xi_2 + i\sin t\xi_2) \\ &= E\cos t\xi_1\cos t\xi_2 + iE\sin t\xi_1\cos t\xi_2 + iE\cos t\xi_1\sin t\xi_2 - E\sin t\xi_1\sin t\xi_2 \end{aligned}$$

由 ξ_1, ξ_2 的独立性及 §2.7 系 1

$$\begin{aligned} f_\eta(t) &= E\cos t\xi_1\cos t\xi_2 + iE\sin t\xi_1 E\cos t\xi_2 + iE\cos t\xi_1 E\sin t\xi_2 - E\sin t\xi_1 E\sin t\xi_2 \\ &= Ee^{it\xi_1}Ee^{it\xi_2} = f_{\xi_1}(t)f_{\xi_2}(t). \end{aligned} \tag{11}$$

对一般的 n, 证明完全类似，或用归纳法证 □

注意： (iv) 的逆是不真的，见例 5.

(v) 设随机变数 ξ 的 n 阶矩存在，则 $f_\xi(t)$ 可微分 $k(k \leqslant n)$ 次，而且

$$f_\xi^{(k)}(0) = i^k E\xi^k \quad (k \leqslant n). \tag{12}$$

证

$$\left|\frac{d^k}{dt^k}(e^{itx})\right| = |i^k x^k e^{itx}| \leqslant |x|^k. \tag{13}$$

根据假定 $\int_{-\infty}^{\infty} |x|^k dF(x) < \infty$, 故下式中在积分号下微分合法 [2] 而有

$$\begin{aligned} f_\xi^{(k)}(t) &= \frac{d^k}{dt^k}f(t) = \frac{d^k}{dt^k}\left[\int_{-\infty}^{\infty} e^{itx}dF(x)\right] \\ &= \int_{-\infty}^{\infty}\left(\frac{d^k}{dt^k}e^{itx}\right)dF(x) = i^k\int_{-\infty}^{\infty} x^k e^{itx}dF(x). \ (k \leqslant n) \end{aligned} \tag{14}$$

在此式中取 $t = 0$ 即得 (12) □

[1] 这性质的反面不成立，见例 5.

[2] 详见 [3] §7.3

(二) 特征函数与分布函数的一一对应 特征函数显然被分布函数唯一决定；现在来证明一个重要的定理，它说明分布函数如何通过特征函数来表达，而且被特征函数唯一决定.

定理 1 (逆转公式) 设分布函数 $F(x)$ 的特征函数为 $f(t)$, 则对任意 $-\infty < x_1 < x_2 < \infty$, 有 [1]

$$\frac{F(x_2+0)+F(x_2-0)}{2}-\frac{F(x_1+0)+F(x_1-0)}{2} = \frac{1}{2\pi}\lim_{l\to\infty}\int_{-l}^{l}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)\,dt; \tag{15}$$

因此，如果 x_1, x_2 是 $F(x)$ 的连续点，则

$$F(x_2)-F(x_1)=\frac{1}{2\pi}\lim_{l\to\infty}\int_{-l}^{l}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)\,dt. \tag{16}$$

证 证明的关键在于利用数学分析 [2] 中的公式

$$\lim_{l\to\infty}\frac{1}{\pi}\int_0^l\frac{\sin\alpha t}{t}\,dt=\begin{cases}\dfrac{1}{2} & (\alpha>0)\\ 0 & (\alpha=0)\\ -\dfrac{1}{2} & (\alpha<0)\end{cases} \tag{17}$$

以及交换几次极限的次序. 由 (1) 得

$$\begin{aligned}I_l&=\frac{1}{2\pi}\int_{-l}^{l}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)\,dt\\&=\frac{1}{2\pi}\int_{-l}^{l}\int_{-\infty}^{\infty}\frac{e^{-itx_1}-e^{-itx_2}}{it}e^{itx}dF(x)\,dt.\end{aligned}$$

利用不等式 [3]：对任意实数 α

$$|e^{i\alpha}-1|\leqslant|\alpha|.$$

可见被积函数的绝对值不超过 x_2-x_1, 由 Fubini 定理可交换积分次序而得

$$\begin{aligned}I_l&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-l}^{l}\frac{e^{it(x-x_1)}-e^{it(x-x_2)}}{it}\,dt\right]dF(x)\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_0^l\frac{e^{it(x-x_1)}-e^{-it(x-x_1)}-e^{it(x-x_2)}+e^{-it(x-x_2)}}{it}\,dt\right]dF(x)\\&=\frac{1}{\pi}\int_{-\infty}^{\infty}\left[\int_0^l\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]dF(x)\\&=\int_{(-\infty,x_1)}+\int_{\{x_1\}}+\int_{(x_1,x_2)}+\int_{\{x_2\}}+\int_{(x_2,\infty)}g(l,x;x_1,x_2)\,dF(x),\end{aligned} \tag{18}$$

[1] 当 $t=0$ 时，按连续性补定义 $\frac{e^{-itx_1}-e^{-itx_2}}{it}$ 之值为 x_2-x_1.

[2] 参看 [34]

[3] 实际上，如 $\alpha\geqslant 0$, 有 $|e^{i\alpha}-1|=\left|\int_0^\alpha e^{ix}\,dx\right|\leqslant\int_0^\alpha|e^{ix}|\,dx=\alpha$, 如 $\alpha<0$, 则 $|e^{i\alpha}-1|=|e^{i\alpha}(e^{i|\alpha|}-1)|=|e^{i|\alpha|}-1|\leqslant|\alpha|$,

其中 $\{y\}$ 表只含一点 y 的单点集，又

$$g(l,x;x_1,x_2)=\frac{1}{\pi}\int_0^l\left[\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right]dt. \tag{19}$$

注意由 (17) 可见 $|g(l,x;x_1,x_2)|$ 有界，由勒贝格控制收敛定理，当 $l\to\infty$ 时，可在 (18) 右方积分号下取极限，故

$$\lim_{l\to\infty}I_l=\int_{(-\infty,x_1)}+\int_{\{x_1\}}+\int_{(x_1,x_2)}+\int_{\{x_2\}}+\int_{(x_2,\infty)}\lim_{l\to\infty}g(l,x;x_1,x_2)\,dF(x).$$

但由 (17)

$$\lim_{l\to\infty}g(l,x;x_1,x_2)=\begin{cases}0, & x\in(-\infty,x_1)\cup(x_2,\infty)\\ \dfrac{1}{2}, & x=x_1\text{或 } x=x_2,\\ 1, & x\in(x_1,x_2).\end{cases}$$

代入上式并以 $F(A)$ $(A\in\mathcal{B}_1)$ 表 $F(x)$ 对应的分布，即得

$$\begin{aligned}\lim_{l\to\infty}I_l&=0+\frac{1}{2}F(\{x_1\})+F((x_1,x_2))+\frac{1}{2}F(\{x_2\})+0\\&=\frac{F(x_1+0)-F(x_1-0)}{2}+F(x_2-0)-F(x_1+0)+\frac{F(x_2+0)-F(x_2-0)}{2}\\&=\frac{F(x_2+0)-F(x_2-0)}{2}-\frac{F(x_1+0)-F(x_1-0)}{2}.\end{aligned}$$

于是 (15) 式得证 □

以 x,y 代替 (16) 中的 x_2 及 x_1, 可见在 $F(x)$ 的每一连续点 x 上，当 y 沿 $F(x)$ 的连续点而趋于 $-\infty$ 时有

$$F(x)=\frac{1}{2\pi}\lim_{y\to\infty}\lim_{l\to\infty}\int_{-l}^{l}\frac{e^{-ity}-e^{-itx}}{it}f(t)\,dt. \tag{20}$$

系 1 (唯一性定理) 二分布函数 $F_1(x)$, $F_2(x)$ 恒等的充要条件是它们各自的特征函数 $f_1(t)$, $f_2(t)$ 恒等.

证 如 $F_1(x)=F_2(x)$, 一切 x, 由 (1) 可见 $f_1(t)=f_2(t)$, 一切 t. 反之，设 $f_1(t)=f_2(t)$, 一切 t, 以 A 表 $F_1(x)$ 及 $F_2(x)$ 的不连续点的集， A 是可列集，由假设及 (20) 可见

$$F_1(x)=F_2(x),\quad x\notin A. \tag{21}$$

今设 $y\in A$, 取一列 $x_n\notin A$, $x_n\downarrow y$. 由分布函数的右连续性

$$F_1(y)=\lim_{n\to\infty}F_1(x_n)=\lim_{n\to\infty}F_2(x_n)=F_2(y)\quad(y\in A).$$

由此及 (21) 知 $F_1(x)$ 与 $F_2(x)$ 恒等 □

系 2 设特征函数 $f(t)$ 绝对可积，即 $\int_{-\infty}^{\infty}|f(t)|\,dt<\infty$, 则对应的分布函数 $F(x)$ 是连

续型的，$F'(x)$ 处处存在，有界而且连续，又对一切 x [1]

$$F'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} f(t)\,dt. \tag{22}$$

证 考虑 $G(x) = \frac{F(x+0)+F(x-0)}{2}$, 如能证 $G'(x)$ 存在，则 $G(x)$ 在点 x 连续，从而 $F(x)$ 也在点 x 连续，于是 $G(x) = F(x)$ 而 $G'(x) = F'(x)$. 故只要对 $G'(x)$ 证明系 2.

由 (15) 得知对 $h > 0$ 有

$$\begin{aligned}\frac{G(x+2h)-G(x)}{2h} &= \frac{1}{2\pi}\lim_{l\to\infty}\int_{-l}^{l}\frac{e^{-itx}-e^{-it(x+2h)}}{2thi}f(t)\,dt\\ &= \frac{1}{2\pi}\lim_{l\to\infty}\int_{-l}^{l}e^{-it(x+h)}\frac{e^{ith}-e^{-ith}}{2^{thi}}f(t)\,dt\\ &= \frac{1}{2\pi}\lim_{l\to\infty}\int_{-l}^{l}e^{-it(x+h)}\frac{\sin th}{th}f(t)\,dt.\end{aligned}$$

由于被积函数的绝对值不超过 $|f(t)|$, 按假定 $|f(t)|$ 在 R_1 可积，故被积函数也在 R_1 可积，$\lim\limits_{l\to\infty}\int_{-l}^{l}$ 可换为 $\int_{-\infty}^{\infty}$, 于是

$$\frac{G(x+2h)-G(x)}{2h} = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-it(x+h)}\frac{\sin th}{th}f(t)\,dt \tag{23}$$

仍然根据上述定理及勒贝格控制收敛定理，当时 $h\to 0+0$, 可在积分号下取极限，故

$$G^{+}(x) = \lim_{h\to 0+0}\frac{G(x+2h)-G(x)}{2h} = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f(t)\,dt, \tag{24}$$

$G^{+}(x)$ 表 $G(x)$ 在点 x 的右导数. 对 $h<0$ 考虑 $\frac{G(x)-G(x+2h)}{-2h}$, 同样可得左导数 $G^{-}(x)$ 也存在而且等于 (24) 中右方值. 这样便证明了：对一切 $x\in R_1$ 有

$$G'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f(t)\,dt. \tag{25}$$

由于 $|e^{-itx}f(x)| = |f(t)|$, 根据假定知 $G'(x)$ 有界，再用一次勒贝格控制收敛定理知 $G'(x)$ 连续. 既然 $G'(x) = F'(x)$, 一切 $x\in R_1$, 故 (22) 成立，而且

$$F(x) = \int_{-\infty}^{x}F'(x)\,dx,$$

从而 $F(x)$ 是连续型的 □

*(三) **半不变量** 设随机变数 ξ 的 n 阶矩 m_n 存在，由 (12) 可把它的特征函数 $f(t)$ 按泰勒 (Taylor) 级数展开而得重要公式

$$f(t) = 1 + \sum_{k=1}^{n}\frac{m_k}{k!}(it)^k + o(t^n). \tag{26}$$

---

[1] 如果 $f(t)$ 取实数值，利用 $e^{-itx} = \cos tx - i\sin tx$, 分别比较 (22) 式双方虚实部分，可见 (22) 式化为 $F'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\cos tx f(t)\,dt$.

注意，$f(t)$ 的绝对可积性并非必要条件，故即使它不成立时，定理结论仍可能正确. 此时，不妨用 (22) 试算，以求概率密度，再用 (1) 验证.

由 $f(0)=1$ 及 $f(t)$ 的连续性，存在 $\delta>0$, 使 $|t|<\delta$ 时，$f(t)\neq 0$, 故函数 $\log f(t)$ 对 $|t|<\delta$ 有意义而且单值 (只考虑对数函数的主值). 由性质 (v), $\log f(t)$ 的前 n 级导数在 0 点存在，令

$$\chi_k=\frac{1}{i^k}\left[\frac{d^k}{dt^k}\log f(t)\right]_{t=0}\qquad (k\leqslant n). \tag{27}$$

称 χ_k 为随机变数 ξ(或它的分布) 的 k 阶半不变量. 显然与 (26) 类似有

$$\log f(t)=\sum_{k=1}^{n}\frac{\chi_k}{k!}(it)^k+o(t^n). \tag{28}$$

为了求出半不变量与矩之间的关系，形式 [1] 地 (不考虑矩的存在及级数的收敛性) 在 (26)(28) 中令 $n\to\infty$ 并利用

$$\log f(t)=\log\left(1+\sum_{k=1}^{\infty}\frac{m_k}{k!}(it)^k\right)=\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(it)^k, \tag{29}$$

$$f(t)=1+\sum_{k=1}^{\infty}\frac{m_k}{k!}(it)^k=e^{\sum\limits_{k=1}^{\infty}\frac{\chi_k}{k!}(it)^k},$$

比较最后一式中 $(it)^k$ 的系数，可见 χ_n 是 $m_1,\cdots,m_n$ 的多项式， m_n 也是 $\chi_1,\cdots,\chi_n$ 的多项式，例如

$$\left.\begin{aligned}&\chi_1=m_1=E\xi,\\&\chi_2=m_2-(m_1)^2=D\xi,\\&\chi_3=m_3-3m_1m_2+2m_1^3,\\&\chi_4=m_4-3m_2^2-4m_1m_3+12m_1^2m_2-6m_1^4,\\&\quad\cdots\cdots.\end{aligned}\right\} \tag{30}$$

反之

$$\left.\begin{aligned}&m_1=\chi_1,\\&m_2=\chi_2+\chi_1^2,\\&m_3=\chi_3+3\chi_1\chi_2+\chi_1^3,\\&m_4=\chi_4+3\chi_2^2+4\chi_1\chi_3+6\chi_1^2\chi_2+\chi_1^4,\\&\quad\cdots.\end{aligned}\right\}. \tag{31}$$

半不变量具有下列简单性质： 1°. 如随机变数 ξ_1,ξ_2 独立，它们的 k 阶半不变量 $\chi_k^{(1)},\chi_k^{(2)}$ 存在，则 $\xi=\xi_1+\xi_2$ 的 k 阶半不变量 χ_k 为

$$\chi_k=\chi_k^{(1)}+\chi_k^{(2)}. \tag{32}$$

证 以 $f(t),f_1(t),f_2(t)$ 分别表 ξ,ξ_1,ξ_2 的特征函数，由 (10) 得

$$\log f(t)=\log f_1(t)+\log f_2(t). \tag{33}$$

[1] 这样形式地做的原因是：它们之间的关系，如果有的话，一般应与形式地求得一致.

由 (29) 即得 (32) □

称 (32) 为半不变量的可加性；注意，对矩 m_k 没有与 (32) 相应的性质，这是它比矩方便之处.

半不变量命名的原因是由于

2°. 在变换

$$\eta = \xi + b \tag{34}$$

下，半不变量 (除一阶外) 不变， b 为常数.

证. 由 (8)

$$f_\eta(t) = e^{itb} f_\xi(t),$$
$$\log f_\eta(t) = itb + \log f_\xi(t).$$

由 (27) 即知 η 及 ξ 的 $k(k > 1)$ 阶半不变量相等，只要它们存在 □

(四) **例.**

例 1 对二项分布 $B(n,p)$,

$$f(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{itk} = \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k q^{n-k} = (pe^{it} + q)^n. \tag{35}$$

设有贝努里实验， A 在每次试验中出现的概率为 p, $q = 1 - p$, 定义随机变数 ξ_j:

$$\xi_j = \begin{cases} 1, & \text{如第 } j \text{ 次试验中 } A \text{ 出现}, \\ 0, & \text{反之}. \end{cases}$$

则 ξ_j $j = 1, \cdots, n$ 独立同分布，分布为 $\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$, 故

$$f_{\xi_j}(t) = pe^{it\cdot 1} + qe^{it\cdot 0} = pe^{it} + q. \tag{36}$$

令 $\eta_n = \sum\limits_{j=1}^{n} \xi_j$, η_n 是前 n 次实验中 A 出现的总次数，由 (iv) 得

$$f_{\eta_n}(t) = (pe^{it} + q)^n. \tag{37}$$

回忆 η_n 有二项分布 $B(n,p)$, 于是重新得到 (35).

设 Y, Z 是二独立随机变数. 分别有二项分布 $B(n,p), B(m,p)$. 于是它们的卷积 $B(n,p) * B(m,p)$ 是 $X = Y + Z$ 的分布. 另方面，由 (iv) 及 (35) 知 X 的特征函数是 $(pe^{it} + q)^{m+n}$. 由系 1 知分布函数被它的特征函数唯一决定；亦即不可能有两个不同的分布函数，它们有相同的特征函数. 既然 $B(n+m,p)$ 的特征函数是 $(pe^{it} + q)^{n+m}$, 故

$$B(n,p) * B(m,p) = B(n+m,p). \tag{38}$$

这证明参数为 n,p 的二项分布与参数为 m,p 的二项分布的卷积也是二项分布. 后者的参数为 $n+m, p$.

从贝努里实验来看，(38) 是明显的，因为 $B(n,p)*B(m,p)$ 是 n 次实验中 A 出现次数与另外 m 次中 A 出现次数的和的分布，也即是 $n+m$ 次中 A 出现次数的分布，故应是 $B(n+m,p)$.

由 (12)(37) 可求得

$$E\eta_n = np,\quad D\eta_n = E\eta_n^2 - (E\eta_n)^2 = npq.$$

这在 §2.9 中已求出．由 (30),η_n 的前二阶半不变量为

$$\chi_1 = np,\quad \chi_2 = npq. \tag{39}$$

这也可直接从 (27) 及 (37) 算得.

例 2 参数为 $\lambda(\lambda>0)$ 的普阿松分布 $P(\lambda)$ 的特征函数为

$$f(t) = \sum_{k=0}^{\infty} e^{itk}\cdot e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!} = e^{-\lambda}\cdot e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}. \tag{40}$$

由此可见

$$P(\lambda)*P(\mu) = P(\lambda+\mu),\ (\lambda>0,\ \mu>0). \tag{41}$$

为求半不变量，由 (40) 得

$$\log f(t) = \lambda(e^{it}-1) = \lambda\sum_{k=1}^{\infty}\frac{(it)^k}{k!}.$$

故根据 (29) 有

$$\chi_k = \lambda,\ (k=1,2,\cdots).$$

例 3 正态分布 $N(0,1)$ 的特征函数为

$$f(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx-\frac{x^2}{2}}\,dx = e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}}\,dx.$$

最后一积分是复变函数 $e^{-\frac{z^2}{2}}$ 在复平面上沿平行于实轴的直线 $y=-it$ 的积分，由闭路积分的理论可见此积分等于同一函数沿实轴的积分 [1]，即为 $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}$, 故

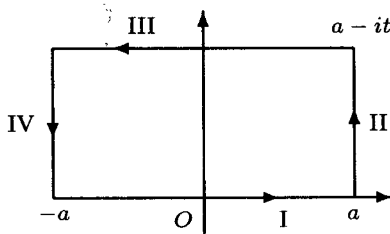
$$f(t) = e^{-\frac{t^2}{2}}. \tag{42}$$

[1]实际上，考虑闭路

$c = \{-a\to a\to a-it\to -a-it\to -a\}$,

它由 I,II,III,IV 四线段构成．由于 $e^{-\frac{z^2}{2}}$ 在 C 中解析，故 $\int_C e^{-\frac{z^2}{2}}\,dz = 0$. 但当 z 位于线段 II,IV 时，令 $z=x+iy$, 则 $\left|e^{-\frac{z^2}{2}}\right| = \left|e^{-\frac{x^2}{2}+\frac{y^2}{2}-ixy}\right| \leqslant e^{-\frac{a^2}{2}+\frac{t^2}{2}}$.

故 $e^{-\frac{z^2}{2}}$ 沿 II,IV 的积分当 $a\to\infty$ 时趋于 0, 从而 $\int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}}\,dx = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}$.



今考虑正态分布 $N(a,\sigma)$, 设它是随机变数 ξ 的分布，那么标准化随机变数 ξ^*

$$\xi^* = \frac{\xi - a}{\sigma} \quad (\sigma > 0)$$

的分布是 $N(0,1)$. 根据 (8)(42) 知 ξ(或 $N(a,\sigma)$) 的特征函数为

$$f_\xi(t) = e^{iat} f_\xi^*(\sigma t) = e^{iat - \frac{(\sigma t)^2}{2}}. \tag{43}$$

由此知

$$N(a_1,\sigma_1) * N(a_2,\sigma_2) = N(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2}). \tag{44}$$

例 4 设随机变数 ξ 在 $[a-h, a+h]$ 上有均匀分布，因而密度为

$$\varphi(x) = \begin{cases} \dfrac{1}{2h}, & x \in [a-h, a+h], \ (h > 0) \\ 0, & x \overline{\in} [a-h, a+h]. \end{cases} \tag{45}$$

(参看 §2.2 例 2), 特征函数为

$$\begin{aligned} f(t) &= \frac{1}{2h} \int_{a-h}^{a+h} e^{itx}\, dx = \frac{1}{2h} \left[\frac{e^{itx}}{it} \right]_{a-h}^{a+h} \\ &= \frac{1}{2h} \frac{e^{it(a+h)} - e^{it(a-h)}}{it} = e^{iat} \frac{\sin th}{th}. \end{aligned} \tag{46}$$

用 (12)(46) 固然可求出 k 阶矩 m_k, 但这里不如直接计算：

$$m_k = \frac{1}{2h} \int_{a-h}^{a+h} x^k\, dx = \frac{1}{2h} \frac{(a+h)^{k+1} - (a-h)^{k+1}}{k+1} \tag{47}$$

作线性变换 $\eta = \frac{\xi - (a-h)}{2h}$, 则易见 η 在 $[0,1]$ 上有均匀分布，密度为

$$\varphi_\eta(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & \overline{\in}[0,1]. \end{cases} \tag{48}$$

均匀分布的重要性还由于下列事实：设随机变数 α 的分布函数 $F(x)$ 是 x 的连续函数，则随机变数 $\beta = F(\alpha)$ 在 $[0,1]$ 上有均匀分布.

实际上，对任意实数 $y \in [0,1]$, 由 $F(x)$ 的连续性，方程

$$F(x) = y \tag{49}$$

至少有一解，记为 $x = F^{-1}(y)$. 于是

$$P(\beta \leqslant y) = P(F(\alpha) \leqslant y) = \begin{cases} 0, & y < 0 \\ P(\alpha \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y, & y \in [0,1] \\ 1, & y > 1. \end{cases}$$

***例 5** 设 (ξ_1, ξ_2) 有联合密度函数为

$$\varphi(x,y) = \left\{ \begin{array}{ll} \dfrac{1}{4}[1 + xy(x^2 - y^2)], & \text{如} |x| \leqslant 1,\ |y| \leqslant 1, \\ 0, & \text{反之}. \end{array} \right\} \tag{50}$$

以 $f_i(t)$ 表 ξ_i 的特征函数， $i=1,2$, 并令

$$\xi=\xi_1+\xi_2$$

的特征函数为 $f(t)$, 我们证明

$$f(t)=f_1(t)f_2(t). \tag{51}$$

然而 ξ_1,ξ_2 不独立. 这说明性质 (iv) 之逆是不正确的. 实际上，ξ_1,ξ_2 的密度函数 $\varphi_1(x),\varphi_2(y)$ 在 $|x|\leqslant 1$ 及 $|y|\leqslant 1$ 时分别为

$$\left.\begin{aligned}\varphi_1(x)&=\int_{-1}^{1}\frac{1}{4}[1+xy(x^2-y^2)]\,dy=\frac{1}{4}\left[y+\frac{x^3y^2}{2}-\frac{xy^4}{4}\right]_{-1}^{1}=\frac{1}{2},\\ \varphi_2(y)&=\int_{-1}^{1}\frac{1}{4}[1+xy(x^2-y^2)]\,dx=\frac{1}{4}\left[x+\frac{x^4y}{4}-\frac{x^2y^3}{2}\right]_{-1}^{1}=\frac{1}{2}.\end{aligned}\right\}$$

故 ξ_1,ξ_2 都在 $[-1,1]$ 中均匀分布. 既然 $\varphi_1(x)\varphi_2(y)=\frac{1}{4}\neq\varphi(x,y)$, $(|x|\leqslant 1,|y|\leqslant 1)$, 故 ξ_1,ξ_2 不独立.

下证 (51). 由 §2.8(二) 的 (4_1) 式，知 ξ 的密度函数 $\psi(z)$ 为

$$\psi(z)=\int_{-\infty}^{\infty}\varphi(x,z-x)\,dx. \tag{52}$$

由于 $\varphi(x,y)$ 在矩形 $|x|\leqslant 1$, $|y|\leqslant 1$ 外为 0, 故对每固定的 z,(52) 中的积分只要同时满足下二不等式

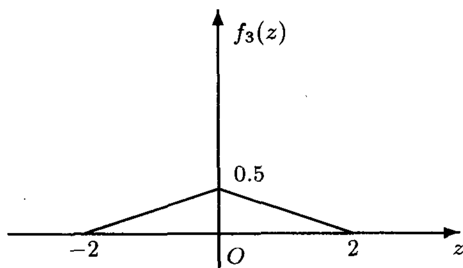
$$|x|\leqslant 1,\quad z-1\leqslant x\leqslant z+1 \tag{53}$$

的 x 进行，当 $-2\leqslant z\leqslant 0$ 时，满足上二不等式的 x 的变化范围为 $[-1,z+1]$; 当 $2\geqslant z\geqslant 0$ 时，范围为 $[z-1,1]$; 当 $|z|>2$ 时，没有 x 能同时满足上二不等式. 故

$$\psi(z)=\left\{\begin{array}{ll}\displaystyle\int_{-1}^{z+1}\frac{1}{4}(1+3z^2x^2-2zx^3-z^3x)dx=\frac{1}{4}(2+z), & \text{如}-2\leqslant z\leqslant 0,\\ \displaystyle\int_{z-1}^{1}\frac{1}{4}(1+3z^2x^2-2zx^3-z^3x)dx=\frac{1}{4}(2-z), & \text{如}0<z\leqslant 2,\\ 0, & \text{如}|z|>2.\end{array}\right\} \tag{54}$$

$\psi(z)$ 属于三角分布的密度函数 (见下图). 由上式

$$\begin{aligned}f(t)&=\frac{1}{4}\left[\int_{-2}^{1}(2+z)e^{itz}\,dz+\int_{0}^{2}(2-z)e^{itz}\,dz\right]\\&=\frac{1}{4}\left[\frac{2-e^{2it}-e^{-2it}}{t^2}\right]\\&=\frac{1}{2t^2}\left(1-\frac{e^{2it}+e^{-2it}}{2}\right)\\&=\frac{1}{2t^2}(1-\cos 2t)=\left(\frac{\sin t}{t}\right)^2;\end{aligned}$$



又同例 4, 得

$$f(t)=\frac{\sin t}{t}\cdot\frac{\sin t}{t}=f_1(t)f_2(t).$$

证明过程中还蕴含下列事实：两个在 $[-1,1]$ 中均匀分布的卷积是一三角分布，后者的密度函数为 (54) 中的 $\psi(z)$.

*§2.12 多元特征函数

(一) 定义与性质. 设已给 n 元分布函数 $F(x_1,\cdots,x_n)$, 它的特征函数 $f(t_1,\cdots,t_n)$, $t_i\in R\ (i=1,\cdots,n)$, 定义为

$$f(t_1,\cdots,t_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i(t_1x_1+\cdots+t_nx_n)}\,dF(x_1,\cdots,x_n). \tag{1}$$

记 $t=(t_1,\cdots,t_n), x=(x_1,\cdots,x_n)$, $t'=\begin{pmatrix}t_1\\ \vdots\\ t_n\end{pmatrix}$, 则上式可简写为

$$f(t)=\int_{R_n}e^{ixt'}\,dF(x). \tag{2}$$

n 元特征函数的理论与一元时类似，故叙述从简.

随机向量 $\xi_1,\cdots,\xi_n)$ 的特征函数 $f(t_1,\cdots,t_n)$ 定义为它的分布函数 $F(x_1,\cdots,x_n)$ 的特征函数，即

$$f(t)=\int_{R_n}e^{ixt'}\,dF(x)=Ee^{i\xi t'}=Ee^{i(t_1\xi_1+\cdots+t_n\xi_n)} \tag{3}$$

(i) 特征函数 $f(t)$ 在 R_n 中均匀连续，而且

$$|f(t_1,\cdots,t_n)|\leqslant f(0,\cdots,0)=1,$$
$$f(-t_1,\cdots,-t_n)=\overline{f(t_1,\cdots,t_n)}.$$

(ii) 设 $F_{j_1,\cdots,j_k}(x_{j_1},\cdots,x_{j_n})$ $(k\leqslant n)$ 是 $F(x_1,\cdots,x_n)$ 对应于 $(j_1,\cdots,j_k)$ 的边沿分布函数，即

$$F_{j_1,\cdots,j_k}(x_{j_1},\cdots,x_{j_k})=F(y_1,\cdots,y_n)\Big|_{\substack{y_i=x_i,\ \text{如}i=j_1,\cdots,j_k,\\ y_i=\infty,\ \text{反之}.}}$$

则 $F_{j_1,\cdots,j_k}(x_{j_1},\cdots,x_{j_k})$ 的特征函数为

$$f_{j_1,\cdots,j_k}(t_{j_1},\cdots,t_{j_k})=f(s_1,\cdots,s_n)\Big|_{\substack{s_i=t_i,\ \text{如}i=j_1,\cdots,j_k,\\ s_i=0,\ \text{反之}.}}$$

例如， $F_1(x_1)=F(x_1,\infty,\cdots,\infty)$ 的特征函数为 $f(t_1,0,\cdots,0)$. 由此可见，如 $(\xi_1,\cdots,\xi_n)$ 的特征函数为 $f(t_1,\cdots,t_n)$, 则 $(\xi_{j_1},\cdots,\xi_{j_k})$ 的特征函数为 $f_{j_1,\cdots,j_k}(t_{j_1},\cdots,t_{j_k})$, ξ_1 的特征函数为 $f(t_1,0,\cdots,0)$.

(iii) 如果 $(\xi_1,\cdots,\xi_n)$ 独立，由 (3) 及刚才所述，有

$$f(t_1,\cdots,t_n)=Ee^{i(t_1\xi_1+\cdots+t_n\xi_n)}=\prod_{j=1}^{n}Ee^{it_j\xi_j}=\prod_{j=1}^{n}f_{\xi_j}(t_j), \tag{4}$$

其中 $f_{\xi_j}(t)$ 表 ξ_j 的特征函数.

(iv) 矩 $E\xi_1^{k_1}\xi_2^{k_2}\cdot\xi_n^{k_n}$ 与特征函数间有关系

$$E\xi_1^{k_1}\xi_2^{k_2}\cdot\xi_n^{k_n}=(i)^{-\sum\limits_{j=1}^{n}k_j}\left[\frac{\partial^{k_1+k_2+\cdots+k_n}f(t_1,t_2,\cdots,t_n)}{\partial t_1^{k_1}\partial t_2^{k_2}\cdots\partial t_n^{k_n}}\right]_{t_1=t_2=\cdots=t_n=0}. \tag{5}$$

(v) 设 $a_j,b_j\ (j=1,2,\cdots,n)$ 为常数，则随机向量 $(\alpha_1\xi_1+b_1,\cdots,a_n\xi_n+b_n)$ 的特征函数为

$$e^{i\left(\sum\limits_{k=1}^{n}b_kt_k\right)}\cdot f(a_1t_1,\cdots,a_nt_n). \tag{6}$$

(iv) 设 a_j 为常数，则 (一维) 随机变数 $\eta=\sum\limits_{j=1}^{n}a_j\xi_j$ 的特征函数 $f_\eta(t)\ (t\in R_1)$ 为

$$f_\eta(t)=Ee^{it\eta}=Ee^{i\left(\sum\limits_{j=1}^{n}a_jt\xi_j\right)}=f(a_1t,\cdots,a_nt)\ \ (\text{由}(3)). \tag{7}$$

(二) n **元特征函数的逆转公式**. 证明虽与一元时相似，但由于它的重要性，我们仍给出证明，为此先引进一概念.

设已给 n 元分布函数 $F(x_1,\cdots,x_n)$, 它所决定的 n 维分布记为 $F(A)$, $A\in\mathcal{B}_n$, 考虑任一 n 维区间:

$$\begin{aligned}(a,b]&=((x_1,\cdots,x_n):\ a_j<x_j\leqslant b_j,\ j=1,\cdots,n), \qquad (8)\\ [a,b]&=((x_1,\cdots,x_n):\ a_j\leqslant x_j\leqslant b_j,\ j=1,\cdots,n),\\ (a,b)&=((x_1,\cdots,x_n):\ a_j<x_j<b_j,\ j=1,\cdots,n).\end{aligned}$$

称 $(a,b]$ 为分布函数 $F(x_1,\cdots,x_n)$ 的连续区间，如果 $F([a,b]\setminus(a,b))=0$; 也就是说 F 在 $(a,b]$ 的边界 $[a,b]\setminus(a,b)$ 上的值等于 0, 显然，如果 $(\xi_1,\cdots,\xi_n)$ 的分布函数为 $F(x_1,\cdots,x_n)$, 而 $(a,b]$ 是后者的连续区间，那么

$$P((\xi_1,\cdots,\xi_n)\in[a,b]\setminus(a,b))=F([a,b]\setminus(a,b))=0. \tag{9}$$

定理 1 (逆转公式) 设 n 元分布函数 $F(x_1,\cdots,x_n)$ 的特征函数为 $f(t_1,\cdots,t_n)$, 对任意由 (8) 定义的区间 $(a,b]$, 如果它是此分布函数的连续区间，则

$$F([a,b])=\frac{1}{(2\pi)^n}\lim_{\substack{l_j\to\infty\\ j=1,\cdots,n}}\int_{-l_1}^{l_1}\cdots\int_{-l_n}^{l_n}\frac{e^{-it_1a_1}-e^{-it_1b_1}}{it_1}\cdots\frac{e^{-it_na_n}-e^{-it_nb_n}}{it_n}\times f(t_1,\cdots,t_n)\,dt_1\cdots dt_n. \tag{10}$$

证 以 (1) 代入 (10) 中的 $f(t_1,\cdots,t_n)$, 于是 (10) 右方积分先是对 $x_1,\cdots,x_n$, 再是对 $t_1,\cdots,t_n$ 的积分. 由于被积函数绝对可积，故由富比尼定理，可改变积分次序，于是右方中的积分化为

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\prod_{j=1}^{n}\int_{-l_j}^{l_j}\frac{e^{it_j(x_j-a_j)}-e^{it_j(x_j-b_j)}}{it_j}dt_j\right]dF(x_1,\cdots,x_n). \tag{11}$$

我们有

$$I_{l_j}(x_j)\equiv\int_{-l_j}^{l_j}\frac{e^{it_j(x_j-a_j)}-e^{it_j(x_j-b_j)}}{it_j}dt_j \tag{12}$$

$$=2\int_0^{l_j}\frac{\sin\,t_j(x_j-a_j)}{it_j}dt_j-2\int_0^{l_j}\frac{\sin\,t_j(x_j-b_j)}{t_j}dt_j. \tag{13}$$

利用 §2.11(17) 式得

$$\lim_{l_j\to\infty}I_{l_j}(x_j)=\begin{cases}2\pi, & \text{如}a_j<x_j<b_j,\\ 0, & \text{如 }x_j<a_j\text{ 或 }b_j<x_j.\end{cases}$$

从而

$$\frac{1}{(2\pi)^n}\lim_{\substack{l_j\to\infty\\ j=1,\cdots,n}}I_{l_1}(x_1)\cdots I_{l_n}(x_n)=\begin{cases}1, & \text{如}(x_1,\cdots,x_n)\in(a,b),\\ 0, & \text{如}(x_1,\cdots,x_n)\in\overline{[a,b]},\end{cases} \tag{14}$$

这里 $\overline{[a,b]}=R_n\setminus[a,b]$. 由 (11),(10) 右方值为

$$\frac{1}{(2\pi)^n}\lim_{\substack{l_j\to\infty\\ j=1,\cdots,n}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}I_{l_1}(x_1)\cdots I_{l_n}(x_n)\,dF(x_1,\cdots,x_n).$$

将 R_n 分为三个互不相交集的和： $R_n=(a,b)\cup\overline{[a,b]}\cup([a,b]\setminus(a,b))$, 上值等于

$$\frac{1}{(2\pi)^n}\lim_{\substack{l_j\to\infty\\ j=1,\cdots,n}}\left[\int\cdots\int_{(a,b)}+\int\cdots\int_{\overline{[a,b]}}+\int\cdots\int_{[a,b]\setminus(a,b)}I_{l_1}(x_1)\cdots I_{l_n}(x_n)\,dF(x_1,\cdots,x_n)\right].$$

由 (13) 及 §2.11(17) 知 $|I_{l_1}(x_1)\cdots I_{l_n}(x_n)|<k$ 有界， k 为某正常数，由勒贝格控制收敛定理，可在积分号下取极限. 由于 (14) 知取极限后第二积分化为 0, 第三积分的绝对值不超过 $kF([a,b]\setminus(a,b))$, 由 $[a,b]$ 是连续区间的假设，它也等于 0, 于是只剩下第一积分，再用 (14) 即知 (10) 右方值为 $F([a,b])$ □

由定理 1 就可仿一元情况而得下定理，我们不一一地去深入它的细节了.

唯一性定理 两个 n 元分布函数 $F_1(x_1,\cdots,x_n)$, $F_2(x_1,\cdots,x_n)$ 恒等的充要条件是他们各自的特征函数 $f_1(t_1,\cdots,t_n)$, $f_2(t_1,\cdots,t_n)$ 恒等.

系 1 以 $f(t_1,\cdots,t_n)$, $f_{\xi_j}(t)$ $(t_i,t\in R_1)$ 分别表 $(\xi_1,\cdots,\xi_n)$ 及 ξ_j 的特征函数, $j=1,\cdots,n$, 则随机变数 $\xi_1,\cdots,\xi_n$ 独立的充要条件是

$$f(t_1,\cdots,t_n)=f_{\xi_1}(t_1)\cdots f_{\xi_n}(t_n). \tag{15}$$

证 必要性见 (4). 反之，设 (15) 成立，$(\xi_1,\cdots,\xi_n)$ 的联合分布函数 $F(x_1,\cdots,x_n)$ 的特征函数按定义是 $f(t_1,\cdots,t_n)$, 以 $F_{\xi_j}(x)$ $(x\in R_1)$ 表 ξ_j 的分布函数，显然

$$G(x_1,\cdots,x_n)=F_{\xi_1}(x_1)\cdots F_{\xi_n}(x_n)$$

也是 n 元分布函数. 然而 $G(x_1,\cdots,x_n)$ 的特征函数为

$$\begin{aligned}&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i(t_1x_1+\cdots+t_nx_n)}\,dF_{\xi_1}(x_1)\cdots F_{\xi_n}(x_n)\\&=\prod_{j=1}^{n}\left[\int_{-\infty}^{\infty}e^{it_jx_j}\,dF_{\xi j}(x_j)\right]=\prod_{j=1}^{n}f_{\xi_j}(t_j)\\&=f(t_1\cdots,t_n)\qquad(\text{由 }(15))\end{aligned}$$

于是得证 $F(x_1,\cdots,x_n)$ 及 $G(x_1,\cdots,x_n)$ 有相同的特征函数，由唯一性定理此二函数恒等，故

$$F(x_1,\cdots,x_n)=F_{\xi_1}(x_1)\cdots F_{\xi_n}(x_n),\tag{16}$$

亦即 $\xi_1,\cdots,\xi_n$ 独立 □

(三) n **维正态分布的性质**. 作为重要的例，试研究 n 维正态分布. 设 n 维随机向量 $\xi=(\xi_1,\cdots,\xi_n)$ 有 n 维正态分布 $N(a,B)$, 密度为

$$f_\xi(x_1\cdots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}e^{-\frac{1}{2}\sum\limits_{j,k=1}^{m}r_{jk}(x_k-a_k)(x_j-a_j)},\tag{17}$$

其中 B 为 n 级正定对称矩阵. $|B|$ 表 B 的行列式的值，$(r_{jk})=B^{-1}$ (参看 §2.6 例 7), 因而 $r_{jk}=\frac{|B_{jk}|}{|B|}$, $|B_{jk}|$ 表行列式 $|B|$ 中元素 b_{jk} 的代数余子式，记 $a=(a_1,\cdots,a_n)$.

引理 1 存在 n 级正交矩阵 T, 使随机向量

$$\eta=(\xi-a)T'\ \ (T'\text{ 为 }T\text{ 的转置矩阵})\tag{18}$$

中的分量 $\eta_1,\cdots,\eta_n$ 是独立随机变数. η_j 有一维正态分布 $N(0,\sqrt{d_i})$, $d_i>0$.

证 考虑 §2.6 例 7 中的正交矩阵 T, 使

$$TBT'=D=\begin{pmatrix}d_1&&&0\\&d_2&&\\&&\ddots&\\0&&&d_n\end{pmatrix}\quad(d_j>0,\ j=1,\cdots,n)\tag{19}$$

由于 T 的行列式的绝对值为 1, 根据 §2.8 定理 1, 知 (18) 所定义的 η 有密度为 [1]

$$f_\eta(y_1,\cdots,y_n)=\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi d_j}}e^{-y_j^2/2d_j}\tag{20}$$

[1] 参看 §2.6 中 (28) 式的证明

由此可推得 η_j 的密度为 $\frac{1}{\sqrt{2\pi d_j}}e^{-y^2/2d_j}$, $(y \in R_1)$, 故

$$f_\eta(y_1, \cdots, y_n) = \prod_{j=1}^n f_{\eta_j}(y_j).$$

从而得证 $(\eta_1, \cdots, \eta_n)$ 的独立性 □

引理 1 说明经过某线性变换后，可把具有分布为 $N(a, B)$ 的随机向量 ξ 变为独立随机向量 η, η 也是正态的，分布为 $N(0, D)$.

定理 2 1° 随机向量 ξ 的特征函数为

$$\varphi_\xi(t_1, \cdots, t_n) = e^{i\sum_{j=1}^n a_j t_j - \frac{1}{2}\sum_{j,k=1}^n b_{jk} t_j t_k}; \tag{21}$$

2° $\xi = (\xi_1, \cdots, \xi_n)$ 的任一子向量 $(\xi_{k_1}, \cdots, \xi_{k_m})$, $(m \leqslant n)$ 也是正态的，分布为 $N(\widetilde{a}, \widetilde{B})$, 其中 $\widetilde{a} = (a_{k_1}, \cdots, a_{k_m})$, $\widetilde{B}$ 为保留 B 中第 $k_1, \cdots, k_m$ 行 (与列), 而抛去其它行 (与列) 的 m 级矩阵.

3°

$$E\xi_j = a_j, \tag{22}$$

$$E(\xi_j - a_j)(\xi_k - a_k) = b_{jk}. \tag{23}$$

4° $\xi_1, \cdots, \xi_n$ 相互独立的充要条件是两两互不相关.

证 1° 考虑 (18) 中的 η 与 T, 注意 $T' = T^{-1}$ (T^{-1} 为 T 的逆矩阵). 由引理 1 及 §2.11(43) 式知 η 的特征函数为

$$\varphi_n(t_1, \cdots, t_n) = e^{-\sum_{j=1}^n \frac{d_j t_j^2}{2}}. \tag{24}$$

采用矩阵符号，由 (18) 得 $\xi = a + \eta T$, 故 ξ 的特征函数为

$$\varphi_\xi(t) = Ee^{i(a+\eta T)t'} = e^{iat'}Ee^{i\eta(Tt')}. \tag{25}$$

作变换

$$Tt' = s' \tag{26}$$

由 (24) 得

$$Ee^{i\eta(Tt')} = Ee^{i\eta s'} = \varphi_\eta(s_1, \cdots, s_n) = e^{-\sum_{j=1}^n \frac{d_j s_j^2}{2}}. \tag{27}$$

另方面，根据二次型理论知：由于 T 满足 (19), 故变换 (26) 把二次型 $t'Bt = \sum_{j,k=1}^n b_{jk} t_j t_k$ 变到 $\sum_{j=1}^n \frac{d_j s_j^2}{2}$; 既然 T^{-1} 存在，由此及 (27) 可见，

$$Ee^{i\eta(Tt')} = e^{-t'Bt}.$$

以此代入 (25) 得

$$\varphi_\xi(t) = e^{iat - t'Bt},$$

此即 (21).

2° 由 (ii) 可见 $(\xi_{k_1},\cdots,\xi_{k_m})$ 的特征函数可自 $\varphi_\xi(t_1,\cdots,t_n)$ 中令 $t_l=0, l\neq k_1,\cdots,l\neq k_m$, 而得, 即

$$e^{i\sum\limits_{u=1}^{m} a_{k_u}t_{k_u}-\frac{1}{2}\sum\limits_{u,v=1}^{m} b_{k_uk_v}t_{k_u}t_{k_v}}$$

然而 $N(\widetilde{a},\widetilde{B})$ 的特征函数也是此函数, 故由唯一性定理即得证 (ii).

3° 由 (ii) 知 ξ_j 有一维正态分布, 故 $E\xi_j$ 及 $D\xi_j$ 存在, 再由 $E|\xi_j-E\xi_j||\xi_k-E\xi_k|\leqslant\sqrt{D\xi_j}\sqrt{D\xi_k}$, 知

$$E(\xi_j-E\xi_j)(\xi_k-E\xi_k)=E\xi_j\xi_k-E\xi_j\cdot E\xi_k$$

存在. 由 (5) 及 (21) 即得 (22)(23).

4° 必要性见 §2.10(二)2°. 反之, 如果 ξ_j,ξ_k 不相关, 即 ξ_j 与 ξ_k 的相关的系数

$$r_{jk}=\frac{E(\xi_j-E\xi_j)(\xi_k-E\xi_k)}{\sqrt{D\xi_j}\sqrt{D\xi_k}}=0,\quad (j\neq k)$$

则 $E(\xi_j-E\xi_j)(\xi_k-E\xi_k)=0$, 从而由 (22)(23) 知 $b_{jk}=0$, 于是由 (21) 得

$$\varphi_\xi(t_1,\cdots,t_n)=\prod_{j=1}^{n} e^{ia_jt_j-\frac{1}{2}b_{jj}t_j^2}=\prod_{j=1}^{n}\varphi_{\xi_j}(t_j), \tag{28}$$

其中 $\varphi_{\xi_j}(t)$ 是 ξ_j 的特征函数. 再由系 1 即得所欲证 □

由定理 2 可见: n 维正态分布 $N(a,B)$ 中, a 是 ξ 的数学期望: $a_i=E\xi_i$, 而 B 是 ξ 的协方差矩阵.

ξ 有 n 维正态分布 $N(a,B)$, 等价于它的特征函数由 (21) 给出.

称 $\xi_1,\cdots,\xi_n$ 的任一线性组合 $\sum\limits_{j=1}^{n}\lambda_j\xi_j+\lambda_0$ 为非平凡的, 如果 $\lambda_1,\cdots,\lambda_n$ 不全为 0, $\lambda_i\in R_1(i=0,1,\cdots,n)$.

定理 3 $(\xi_1,\cdots,\xi_n)$ 有 n 维正态分布的充要条件是它的任一非平凡的线性组合 $\eta=\sum\limits_{j=1}\lambda_j\xi_j+\lambda_0$ 有一维正态分布.

证 充分性: 因 ξ_j 也可看成 $(\xi_1,\cdots,\xi_n)$ 的非平凡线性组合 (取 $\lambda_j=1$, 其余 $\lambda_k=0$), 故由假定 ξ_j 有一维正态分布, 从而 $E\xi_j=a_j$, $D\xi_j=b_{jj}$ 存在 (一切 j), 于是

$$b_{jk}=E(\xi_j-a_j)(\xi_k-a_k)$$

也存在. 由此立知: 对非平凡的线性组合 $\eta=\sum\limits_{j=1}^{n}\lambda_j\xi_j$, 有

$$E\eta=\sum_{j=1}^{n}\lambda_ja_j,\quad D^2\eta=\sum_{j,k=1}^{n}b_{jk}\lambda_j\lambda_k.$$

根据假定 η 有一维正态分布, 故 η 的分布为 $N\Big(\sum\limits_{j=1}^{n}\lambda_ja_j,\sqrt{\sum\limits_{j,k=1}^{n}b_{jk}\lambda_j\lambda_k}\Big)$, η 的特征函数为

$$Ee^{it\eta}=\exp\left\{it\sum_{i=1}^{n}\lambda_ja_j-\frac{t^2}{2}\sum_{j,k=1}^{n}b_{jk}\lambda_j\lambda_k\right\}.$$

令 $t=1$, 得

$$Ee^{i(\lambda_1\xi_1+\cdots+\lambda_n\xi_n)}=\exp\left\{i\sum_{j=1}^{n}\lambda_j a_j-\frac{1}{2}\sum_{j,k=1}^{n}b_{jk}\lambda_j\lambda_k\right\}.$$

由于 $\lambda_j\in R_1$ 可任意，故把 $(\lambda_1,\cdots,\lambda_n)$ 看成 R_n 中的变点后，上式便表明 $(\xi_1,\cdots,\xi_n)$ 有 n 维正态分布.

必要性：设 ξ 的特征函数由 (21) 给出，考虑它的任一非平凡线性组合 $\eta=\sum\limits_{j=1}^{n}\lambda_j\xi_j+\lambda_0$ 则

$$\begin{aligned}E(e^{it\eta})&=E\left(e^{i\lambda_0 t}e^{i\sum\limits_{j=1}^{n}\lambda_j t\xi_j}\right)=e^{i\lambda_0 t}E\left(e^{i\sum\limits_{j=1}^{n}(\lambda_j t)\xi_j}\right)\\&=e^{i\lambda_0 t}\exp\left\{i\sum_{j=1}^{n}a_j(\lambda_j t)-\frac{1}{2}\sum_{j,k=1}^{n}b_{jk}(\lambda_j t)(\lambda_k t)\right\}\quad(\text{由}(21))\\&=\exp\left\{i\left(\lambda_0+\sum_{j=1}^{n}a_j\lambda_j\right)t-\frac{t^2}{2}\sum_{j,k=1}^{n}b_{jk}\lambda_j\lambda_k\right\},\end{aligned}$$

这说明 η 有 $N\left(\lambda_0+\sum\limits_{j=1}^{n}a_j\lambda_j,\sqrt{\sum\limits_{j,k=1}^{n}b_{jk}\lambda_j\lambda_k}\right)$ 分布 □

下定理表明正态性在线性变换下保持不变.

定理 4 设 $(\xi_1,\cdots,\xi_n)$ 有 n 维正态分布 $N(a,B)$, 令

$$\eta_l=\sum_{j=1}^{n}\lambda_{lj}\xi_j\quad(l=1,\cdots,r,r\leqslant n;\ \lambda_{lj}\text{常数}).\tag{29}$$

如果 r 级矩阵 (μ_{lh}) 正定，其中

$$\mu_{lh}=\sum_{j,k=1}^{n}\lambda_{lj}\lambda_{hk}b_{jk},$$

则随机向量 $(\eta_1,\cdots,\eta_r)$ 有 r 维正态分布.

证

$$\begin{aligned}E(\eta_l-E\eta_l)(\eta_h-E\eta_h)&=E\left(\sum_{j=1}^{n}\lambda_{lj}(\xi_j-E\xi_j)\sum_{k=1}^{n}\lambda_{hk}(\xi_k-E\xi_k)\right)\\&=\sum_{j,k=1}^{n}\lambda_{lj}\lambda_{hk}E(\xi_j-E\xi_j)(\xi_k-E\xi_k)=\sum_{j,k=1}^{n}\lambda_{lj}\lambda_{hk}b_{jk}=\mu_{lh}.\end{aligned}\tag{30}$$

以 $\varphi_\eta(t_1,\cdots,t_r)$ 表 η 的特征函数，则

$$\begin{aligned}\varphi_\eta(t_1,\cdots,t_r) &= E\left[\exp\left(i\sum_{l=1}^{r} t_l\eta_l\right)\right] \\ &= E\left[\exp\left(it_1\sum_{j=1}^{n}\lambda_{1j}\xi_j+\cdots+it_r\sum_{j=1}^{n}\lambda_{rj}\xi_j\right)\right] \\ &= E\left[\exp\left(i\sum_{j=1}^{n}v_j\xi_j\right)\right],\end{aligned}$$

其中 $v_j=t_1\lambda_{1j}+\cdots+t_r\lambda_{rj}$ $(j=1,\cdots,n)$, 由 (21) 得

$$\begin{aligned}\varphi_\eta(t_1,\cdots,t_r) &= \exp\left(i\sum_{j=1}^{n}a_jv_j-\frac{1}{2}\sum_{j,k=1}^{n}b_{jk}v_jv_k\right) \\ &= \exp\left(i\sum_{j=1}^{n}a_j\sum_{l=1}^{r}t_l\lambda_{lj}-\frac{1}{2}\sum_{j,k=1}^{n}b_{jk}\sum_{l,h=1}^{r}t_lt_h\lambda_{lj}\lambda_{hk}\right) \\ &= \exp\left[i\sum_{l=1}^{r}t_l\left(\sum_{h=1}^{n}a_h\lambda_{lh}\right)-\frac{1}{2}\sum_{l,h=1}^{r}t_lt_h\sum_{j,k=1}^{n}\lambda_{lj}\lambda_{hk}b_{jk}\right] \\ &= \exp\left[i\sum_{l=1}^{r}t_lc_l-\frac{1}{2}\sum_{l,h=1}^{r}\mu_{lh}t_lt_h\right],\end{aligned}$$

其中 $c_l=\sum\limits_{h=1}^{n}a_h\lambda_{lh}$. 由 (30) 知矩阵 (μ_{lh}) 对称，非负定，根据假定它还是正定的，故从上式知 $(\eta_1,\cdots,\eta_r)$ 有 r 维正态分布为 $N(c,(\mu_{lh}))$, $c=(c_1,\cdots,c_r)$ □

以上我们都假定了 B 是正定的. 如果它是非负定但不是正定因而 B^{-1} 不存在时，我们便得到奇异的正态分布. 例如对一维正态分布 $N(a,\sigma)$, 如 $\sigma=0$, 便是一奇异分布，它的质量集中在一个点 a 上. 一般地，设 B 的秩为 $r(<n)$, 可以证明，按这个奇异正态分布而分布的质量，以概率 1 集中在某一个 r 维集合上 (参看 [3],24.3).

*§2.13 若 干 补 充

(一) **概率论与测度论的关系**. 到现在为止，我们已叙述完概率论中的基本概念与基本知识. 在第一章里，从已给的随机试验出发，对此随机试验作若干假设 (例如古典型，几何型等) 后，可以计算某些事件的概率，所用的方法基本上是排列组合，差分方程等. 这一章的内容，可以看成为产生概率论的历史背景，也是概率论发展史中古典阶段 (十九世纪中叶以前) 研究内容的一个缩影.

随机变数的产生是概率论发展过程中的一个飞跃. 由于它的出现，使得概率论研究的对象从个别事件扩大为全面地刻画随机试验的变数，同时，又由于对象的扩大而必须引入其它新的概念与工具如分布，分布函数等等，这样便全面地刷新了概率论的内容. 然而为了使事件，随机变数等基本概念获得严格的数学意义，让它们不至于停留在感性的阶段，还有待于公理系统的建立. 目前为大多数人所接受的公理系统由柯尔莫哥洛夫 (A. H.

Колмогоров) 所建成，这在前面两章内已经看到了．根据这一公理系统，概率论与实变函数论有许多相似之处．其实概率论的某些部分，可以看成测度论 — 后者是实变函数论中一部分内容即度量理论的一般化 — 的一个分支，概率论中许多基本概念都能翻译成测度论中的名词．试给出一张对照表 (对不熟悉测度论的读者，可设想右方栏中的 Ω 为 $[0,1]$,$\mathcal{F}$ 为 $[0,1]$ 中全体 Borel 子集， $P(A)$ 为 A 的勒贝格测度， $A\in\mathcal{F}$; 从而可看出概率论与实变函数论中基本概念的类似性).

	概　　率　　论	测　　度　　论
1	概率空间 Ω	(全测度为 1 的) 测度空间 Ω
2	基本事件 ω	Ω 中的点 ω
3	事件集 $\mathcal{F}$	σ 代数 $\mathcal{F}$
4	事件 A	$\mathcal{F}$ 中的集 A
5	事件的概率 $P(A)$	A 的测度 $P(A)$
6	随机变数 $\xi(\omega)$	关于 $\mathcal{F}$ 可测的函数
7	数学期望 $E\xi(\omega)$	积分 $\int_\Omega \xi(\omega)P(d\omega)$

然而不能把测度论完全代替概率论，这不仅因为概率论的历史较久，而且还主要是由于概率论有它自己侧重研究的方面：随机变数 (或随机向量) 的分布及与分布有关的问题．

伴随着随机变数 ξ 而出现的概念是分布函数 $F_\xi(x)$, 它全面刻划了随机变数的概率性质．至于较粗地刻划 ξ(或 $F_\xi(x)$) 的某些性质的则有 $E\xi$, $D\xi$ 等等．然而在一些问题中，例如求独立随机变数的和的分布时，用分布函数就显得很复杂，于是引进了比较容易运算的特征函数 $f_\xi(t)$, 它与 $F_\xi(x)$ 相互唯一决定．因而理论上 $F_\xi(x)$ 与 $f_\xi(t)$ 处于平等的地位，不过 $f_\xi(t)$ 的概率意义不明显．关于 ξ 的一些概率问题既可用 $F_\xi(x)$, 也可用 $f_\xi(t)$ 来解决，例如求矩问题，求和的分布问题等就是如此．至于什么时候用 $F_\xi(x)$ 还是用 $f_\xi(t)$ 较方便，那就依赖于问题的性质了．

进一步是讨论某些具体分布的性质，问题不在于造许多分布，而是要挑出一些有理论或实际应用价值的分布，并研究它们的性质．在 §2.5 后面附有常用的分布表，以备查用．

(二) 几个有关问题．下面叙述一些与第二章的内容有关的问题，我们只限于提出问题或指明结果而不给出证明，因为它们大都超出本书的范围．

甲．关于分布的研究

1° 一个古典的结果是分布函数 $F(x)(x\in R_1)$ 的分解．在 §2.2 中已看到三种不同的分布函数：连续型的分布函数 $F_1(x)$, 离散型的 $F_2(x)$, 奇异型的 $F_3(x)$. 容易看出：如果 a_1,a_2,a_3 是三个非负常数， $\sum\limits_{i=1}^{3}a_i=1$, 那么

$$F(x)=a_1F_1(x)+a_2F_2(x)+a_3F_3(x) \tag{1}$$

也是一分布函数 (证明与 §2.2(20) 的证同). 有趣的是反面的结论也正确： “任一分布函数可唯一地展为 (1) 的形式，其中 $F_i(x)$, $a_i(i=1,2,3)$ 满足上述条件”. 证明可见参考书 [1, 第九章， §7] 或 [2, 第四章， §11, 178 页), 由此知上述三种分布函数是最基本的，其它的不过是这三种的线性组合．

2° 现在来看一个著名的问题 — 矩问题：设已给一列常数 $\{a_n\}$, 试问在什么条件下存在一分布函数 $F(x)$, 使 $F(x)$ 的 n 阶矩恰为 a_n, $n=1,2,\cdots$? 这样的 $F(x)$ 如果有的话，又是否唯一？

关于这些问题的研究已有专门的书 (例如 [4]). 这里只限于给出唯一性的一个便于应用的结果.

唯一性的问题也可以这样叙述：在什么条件下，分布函数 $F(x)$ 被它的各阶矩 a_n $n=1,2,\cdots$ 所唯一决定.

我们已看到，二项分布，普阿松分布，正态分布被它们的一、二阶矩所唯一决定.

然而，这问题的答案一般却是否定的：不难找到两个不同的分布函数，它们有相同的各阶矩. 例如 [1], 设连续型的 $F_1(x), F_2(x)$ 分别有密度为

$$f_1(x)=\begin{cases} c\exp[-x^u\cos u\pi], & \text{如} x>0,\\ 0, & \text{如} x\leqslant 0.\end{cases}$$

$$f_2(x)=\begin{cases} c[1+\sin(x^u\sin u\pi)]\exp[-x^u\cos u\pi], & \text{如} x>0,\\ 0, & \text{如} x\leqslant 0.\end{cases}$$

其中 $0<u<\frac{1}{2}$, $c=\frac{u(\cos u\pi)^{1/u}}{\Gamma(1/u)}$. 可以证明，它们给出不同的分布函数，然而却有相同的各阶矩：

$$\int_0^\infty x^n f_1(x)\,dx=\frac{\Gamma\left(\frac{n+1}{u}\right)}{\Gamma\left(\frac{1}{u}\right)}(\cos u\pi)^{-n/u}=\int_0^\infty x^n f_2(x)\,dx,\ (n=0,1,2,\cdots). \tag{2}$$

(参看 [5,§1.4]).

然而在某些条件下，矩问题的解唯一，我们有结果

(A) 设 $a_0=1,a_1,a_2,\cdots$ 是某分布函数 $F(x)$ 的各阶矩，它们都有穷，如果级数 $\sum\limits_{n=0}^{\infty}\frac{a_n}{n!}r^n$ 对某 $r>0$ 绝对收敛，则此 $F(x)$ 是唯一的以 a_n 为 n 阶矩 $(n=0,1,2,\cdots)$ 的分布函数 (证见 [3,§15.4] 或 [6, §5.5]).

由 (A) 立刻推出结果

(B) 如果随机变数 ξ 以概率 1 有界，即存在常数 $c>0$ 使 $P(|\xi|<c)=1$, 则 ξ 的分布函数 $F(x)$ 被它的各阶矩 $\{a_n\}$ 唯一决定.

事实上，这时 $F(-c)=0$, $F(c)=1$, 故

$$|a_n|\leqslant\int_{-c}^{c}|x|^nF(dx)\leqslant c^n,$$

$$\sum_{n=0}^{\infty}\left|\frac{a_nc^n}{n!}\right|\leqslant\sum_{n=0}^{\infty}\frac{c^{2n}}{n!}=e^{c^2}<\infty. \tag{3}$$

3° 近年来关于分布的分解问题有不少研究，在 §2.11 例 3 中已经看到，两个具有正态分布的独立随机变数 ξ_1,ξ_2 的和 $\xi_1+\xi_2$ 也有正态分布；或者说，二正态分布的卷积也是正态的. 重要的是反面的结论也成立. 1934 年列威 (P. L'evy) 预言：如果两个独立随机变数的和有正态分布，那么每个随机变数都有正态分布，这预言为克拉美 (H. Gramer) 所证

[1] 较简单的例见 [31] 106 页. 又在 [15] 卷二 (1971 年版)227 页上证明了对数正态分布不被它的各阶矩所唯一决定

明 (见 [14, 定理 18],[12,§6,302 页],[15,169 页] 或 [13, 定理 6.31]). 同样结论对普阿松分布也成立，这由拉依科夫 (Д. А. Райков) 证明 (见 [13, 定理 6.61]). 后来林尼克 (Линник) 在这方面作了不少的工作，他研究了比较一般的分布的分解问题，详细的叙述见他的书 [13].

乙. 关于特征函数研究.

特征函数的研究已积累了大量的文献，许多结果已综合在书 [5][13] 中，研究的问题很多，例如：函数 $f(t)$ $(t \in R_1)$ 是分布函数 (或某种分布函数) 的特征函数的充要条件 (其中包括我们以后要叙述的波赫纳尔 - 辛钦定理), 一般的特征函数的性质；某类特殊的特征函数的性质，譬如说，解析特征函数、周期特征函数的性质，等等.

我们不去叙述具体的结果.

特征函数的研究中积累了许多有意义的例子，试举出一些，这对加深对特征函数的理解无疑是有益的.

例 1 这例说明当 n 为奇数时， §2.11 中特征函数的性质 (v) 的逆是不成立的 [1]：特征函数在 0 点可微分 k 次，但 k 阶矩未必存在.

首先注意一般事实：如密度函数 $\varphi(x)$ 关于原点对称，即如 $\varphi(x) = \varphi(-x)$, 则它对应的特征函数

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\varphi(x)\,dx = 2\int_0^{\infty} \cos tx\varphi(x)\,dx, \tag{4}$$

因而 $f(t)$ 取实值， $-1 \leqslant f(t) \leqslant 1$.

实际上，利用对称性质及 $e^{ix} = \cos x + i\sin x$, 立得

$$\begin{aligned} f(t) &= \int_{-\infty}^{0} e^{itx}\varphi(x)\,dx + \int_0^{\infty} e^{itx}\varphi(x)\,dx = -\int_{\infty}^{0} e^{-itx}\varphi(x)\,dx + \int_0^{\infty} e^{itx}\varphi(x)\,dx \\ &= \int_0^{\infty} (e^{-itx} + e^{itx})\varphi(x)\,dx = 2\int_0^{\infty} \cos tx\varphi(x)\,dx. \end{aligned}$$

今取

$$\varphi(x) = \begin{cases} 0, & \text{如}|x| \leqslant 2, \\ \dfrac{c}{x^2 \log |x|}, & \text{如}|x| > 2, \end{cases} \tag{5}$$

其中 c 为常数，由条件 $\int_{-\infty}^{\infty} \varphi(x)\,dx = 1$ 决定. 于是 $\varphi(x)$ 是关于原点对称的密度函数. 由 (4), 特征函数为

$$f(t) = 2c\int_2^{\infty} \frac{\cos tx}{x^2 \log x}\,dx. \tag{6}$$

由此知 $\frac{1-f(t)}{2c}$ 是 t 的非负，实值偶函数. 利用 $0 \leqslant 1 - \cos x \leqslant \min(2, x^2)$, 对 $t < \frac{1}{2}$ 得

$$\begin{aligned} 0 \leqslant \frac{1-f(t)}{2c} &= \int_2^{\infty} \frac{1-\cos tx}{x^2 \log x}\,dx = \int_2^{1/t} \frac{1-\cos tx}{x^2 \log x}\,dx + \int_{1/t}^{\infty} \frac{1-\cos tx}{x^2 \log x}\,dx \\ &\leqslant 2t^2\int_2^{1/t} \frac{dx}{\log x} + 2\int_{1/t}^{\infty} \frac{dx}{x^2 \log x} = o(-t/\log t) = o(t), \ (t \to 0) \end{aligned}$$

注意 $f(0) = 1$, 故上式说明 $f(t)$ 在 $t = 0$ 有导数为 0.

---

[1] 但对偶数的 n, 逆结论成立，见 [3,§10,1] 或 [4, 第四章， §12, 199 页]

然而由

$$\int_2^a x\varphi(x)\,dx=\int_2^a \frac{c\,dx}{x\log x}=c[\log\log a-\log\log 2]\to\infty\ (a\to\infty)$$

知一阶矩不存在.

例 2 称特征函数 $f(t)$ 为无穷可分的，如对任一正整数 n, 存在一特征函数 $f_n(t)$, 使

$$f(t)=[f_n(t)]^n. \tag{7}$$

无穷可分特征函数所对应的分布叫无穷可分分布；显然，一个分布无穷可分的充要条件是：对任意正整数 n, 它是某 n 个独立同分布的随机变数的和的分布.

试举一些无穷可分分布的例：

(a) 正态分布：

$$f(t)=\exp\left(iat-\frac{\sigma^2t^2}{2}\right),\quad f_n(t)=\exp\left(i\frac{a}{n}t-\frac{\sigma^2t^2}{2n}\right);$$

(b) 普阿松分布：

$$f(t)=\exp(\{\lambda(e^{it}-1)\},\quad f_n(t)=\exp\left\{\frac{\lambda}{n}(e^{it}-1)\right\};$$

(c) 柯西分布：

$$f(t)=\exp\{iut-\lambda|t|\},\quad f_n(t)=\exp\left\{i\frac{u}{n}t-\frac{\lambda}{n}|t|\right\};$$

(d) 单点分布：

$$f(t)=e^{ict},\quad f_n(t)=e^{i\frac{c}{n}t};$$

等等.

无穷可分分布在独立随机变数和的极限理论中起着十分重要的作用，关于这方面的一个初步介绍见 [9, 第九章], 进一步见 [11] 或 [2].

例 3 一特征函数在有限区间内的值不足以唯一决定此特征函数，因而也不足以唯一决定分布函数. 现在来举这样的例. 令

$$f_1(t)=\begin{cases}1-|t|, & \text{当}|t|\leqslant 1,\\ 0, & \text{当}|t|>1.\end{cases} \tag{8}$$

由于 $f_1(t)$ 在 R_1 上绝对可积，故对应的密度函数为

$$\begin{aligned}\varphi_1(x)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f_1(x)\,dt\\&=\frac{1}{2\pi}\int_{-1}^{0}(1+t)e^{-itx}\,dt+\frac{1}{2\pi}\int_0^1(1-t)e^{-itx}\,dt\\&=\frac{1}{2\pi}\left[-\frac{1}{ix}-\frac{1}{(ix)^2}(1-e^{ix})\right]+\frac{1}{2\pi}\left[\frac{1}{ix}-\frac{1}{(ix)^2}(e^{-ix}-1)\right]\\&=\frac{1}{\pi x^2}\left(1-\frac{e^{ix}+e^{-ix}}{2}\right)=\frac{1-\cos x}{\pi x^2}.\end{aligned} \tag{9}$$

现在考虑一离散型分布，它的密度矩阵为

$$\begin{pmatrix} 0 & (2k-1)\pi & k=0,\pm 1,\pm 2,\cdots \\ \frac{1}{2} & \frac{2}{(2k-1)^2\pi^2} & \end{pmatrix}. \tag{10}$$

这分布的特征函数为

$$\begin{aligned} f_2(t) &= \frac{1}{2} + \sum_{k=-\infty}^{\infty} \frac{2}{(2k-1)^2\pi^2} e^{it(2k-1)\pi} \\ &= \frac{1}{2} + \frac{2}{\pi^2} \sum_{k=1}^{\infty} \frac{\cos(2k-1)t\pi + i\sin(2k-1)t\pi}{(2k-1)^2} \\ &= \frac{1}{2} + \frac{4}{\pi^2} \sum_{k=-\infty}^{\infty} \frac{\cos(2k-1)t\pi}{(2k-1)^2}. \end{aligned} \tag{11}$$

今证对 $|t| \leqslant 1$, 有 $f_1(t) = f_2(t)$. 实际上，在区间 $|t| \leqslant 1$ 上把函数 $g(t) = |t|$ 展成傅立叶级数，

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\pi t, \tag{12}$$

其中系数

$$\frac{a_0}{2} = \int_0^1 t\,dt = \frac{1}{2},$$

$$\begin{aligned} a_n &= 2\int_0^1 t\cos n\pi t\,dt = \left[\frac{2t\sin \pi t}{n\pi}\right]_0^1 - \frac{2}{n\pi}\int_0^1 \sin n\pi t\,dt \\ &= -\frac{2}{n\pi}\left[\frac{-\cos n\pi t}{n\pi}\right]_0^1 = 2\frac{\cos n\pi - 1}{n^2\pi^2}, \end{aligned}$$

因而 $a_{2k} = 0,\ (k \neq 0),\ a_{2k-1} = -\frac{4}{(2k-1)^2\pi^2}$. 代入 (12) 得

$$g(t) = |t| = \frac{1}{2} - \frac{4}{\pi^2} \sum_{k=1}^{\infty} \frac{\cos(2k-1)\pi t}{(2k-1)^2}. \tag{13}$$

代入 (8) 后，将所得结果与 (11) 比较，可见

$$f_1(t) = f_2(t), \quad |t| \leqslant 1.$$

然而 $f_1(t), f_2(t)$ 是二不同分布 (一是连续型的，一是离散型的) 的特征函数，故由唯一性定理知 $f_1(t)$ 与 $f_2(t)$ 是二不同的特征函数.

这例由格涅坚科 (Гнеденко) 造出，我们并附带地看到了一个具有周期为 2 的特征函数 $f_2(t)$. 由周期性 $f_2(t)$ 在 $|t| > 1$ 中不会恒等于 0, 因而由此及 (8) 也可看出 $f_1(t)$ 与 $f_2(t)$ 是不同的特征函数.

习 题

1. 试画出下列随机变数的分布函数

$$\text{(i) } \xi : \begin{pmatrix} 0 & \frac{\pi}{2} & \pi \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}; \quad \text{(ii) } \eta = \frac{2}{3}\xi + 2; \quad \text{(iii) } \rho = \cos\xi.$$

2. 设随机变数 ξ 有连续型分布函数 $F(x) = \int_{-\infty}^{x} p(t)\,dt$. 试求下列随机变数的分布函数和密度

$$(i)\ \eta = \xi^{-1}; \quad (ii)\ \eta = |\xi|.$$

3. 设随机变数 ξ 具有连续的分布函数 $F(x)$, 问 (i)$\eta = F(\xi)$ 是不是随机变数？为什么？ (ii) 如果是的话；它的分布函数是什么？分布是什么？

4. 随机变数的密度是 $f(x) = \begin{cases} a\cos x, & -\dfrac{\pi}{2} \leqslant x < \dfrac{\pi}{2}, \\ 0 & x > \dfrac{\pi}{2},\ x < \dfrac{-\pi}{2}. \end{cases}$

(i) 求系数 $a =$?

(ii) 作出 $f(x)$ 以及分布函数的图形.

5. 设 $\xi_1 : \binom{0}{1}$, 而 ξ_2 是任意的随机变数，试证 ξ_1, ξ_2 独立.

6. 求证任何分布函数具有下列性质

$$\lim_{x\to\infty} x\int_{x}^{\infty} \frac{1}{z}\,dF(z) = 0; \quad \lim_{x\to+0} x\int_{x}^{\infty} \frac{1}{z}\,dF(z) = 0;$$
$$\lim_{x\to-\infty} x\int_{-\infty}^{x} \frac{1}{z}\,dF(z) = 0; \quad \lim_{x\to-0} x\int_{-\infty}^{x} \frac{1}{z}\,dF(z) = 0.$$

7. 设二维随机变数 (x_1, x_2) 的联合分布密度是

$$\frac{1}{\Gamma(k_1)\Gamma(k_2)} x_1^{k_1-1}(x_2 - x_1)^{k_2-1} e^{-x_2}, \quad 0 < x_1 \leqslant x_2 < \infty,\ k_1 > 0, k_2 > 0,$$

试求 x_1 和 x_2 的边沿密度.

8. 一本 500 页的书，共有 500 个错字，每个错字等可能地出现在每一页上，试求在给定的一页上至少有三个错字的概率.

9. 在半径为 R, 中心在坐标原点的圆周上任意抛掷一个点 [即所抛点的极角均匀分布在区间 $(-\pi, \pi)$ 内], 试求连结所抛点与 $(-R, 0)$ 的弦长的分布密度.

10. 在 $(0, a)$ 线段上任意投掷两个点， [即其横坐标均匀分布于线段 $(0, a)$ 内] 试求：两点间距离的分布函数.

11. 在线段 $(0, a)$ 上任意投掷 n 个点，设诸点的布列是任意的 [即每点和其他各点独立布列着，并且在 $(0, a)$ 内均匀分布], 试求

(i) 左边第 k 点的横坐标的分布密度；

(ii) 左边第 k 点和第 m 点的横坐标的联合分布密度 $(k < m)$.

12. 设 $\xi_1, \cdots, \xi_k$ 是独立随机变数，有同样的分布函数 $F(x)$. 令

$$U = \min(\xi_1, \cdots, \xi_k), \quad V = \max(\xi_1, \cdots, \xi_k).$$

试证 (U,V) 的联合分布函数是

$$[F(v)]^k - g(u,v)[F(v)-F(u)]^k,$$

其中 $g(u,v)=\begin{cases}1, & u<v,\\ 0, & u\geqslant v.\end{cases}$

13. 设 ξ,η 独立，都服从正态分布 $N(0,1)$; 试求其积的分布密度.

14. 设 ξ,η 独立，分布密度是

(i) $p_\xi(x)=p_\eta(x)=\begin{cases}0, & x\leqslant 0,\\ ae^{-\alpha x}, & x>0(\alpha>0),\end{cases}$

(ii) $p_\xi(x)=p_\eta(x)=\begin{cases}0, & x\leqslant 0, \text{或} x>a,\\ \dfrac{1}{a}, & 0<x\leqslant a.\end{cases}$

试求 $\zeta=\frac{\xi}{\eta}$ 的分布密度.

15. 试求独立随机变数 ξ_1 与 ξ_2 的和的分布密度. 假定

(i) ξ_1,ξ_2 的分布函数是 $F_1(x)=F_2(x)=\frac{1}{2}+\frac{1}{\pi}\operatorname{arctg} x$;

(ii) ξ_1,ξ_2 分别有 $(-5,1)$ 与 $(1,5)$ 内的均匀分布;

(iii) ξ_1,ξ_2 的分布密度为 $p_1(x)=p_2(x)=\frac{1}{2\alpha}e^{-\frac{|x|}{\alpha}}$, $\alpha>0$.

16. 求证：如果 ξ 与 η 独立，且分别有自由度为 m 和 n 的 χ^2- 分布，则

$$\xi+\eta \text{ 与 } \frac{\xi}{\eta} \text{ 也独立.}$$

17. 求证：如果 ξ 与 η 独立，且都有正态分布 $N(0,1)$, 则

$$U=\xi^2+\eta^2,\ \ V=\frac{\xi}{\eta} \text{ 也独立.}$$

18. 设 ξ,η 独立，且有相同分布密度 $p(x)=(\ln\theta)\theta^{-x}$, $0<x<\infty$, $1<\theta<\infty$, 问 $\xi+\eta$ 与 η 是否独立?

19. 设独立随机变数 ξ,η 有相同分布密度 $p(x)=\begin{cases}e^{-x}, & 0<x<\infty,\\ 0, & \text{其他}.\end{cases}$ 问 $\xi+\eta$ 与 $\frac{\xi}{\xi+\eta}$ 是否独立?

20. 对于二项分布，若令 $B(k:n,p)=\sum\limits_{\nu=0}^{k}b(\nu,n,p)$, 其中 $b(\nu,n,p)=\binom{n}{\nu}p^\nu q^{n-\nu}$. 试证：

$$B(k;n+1,p)=B(k;n,p)-pb(k;n,p),$$

$$B(k+1;n+1,p)=B(k;n,p)+qb(k+1;n,p).$$

21. 求证：对 $0\leqslant p\leqslant 1$, $q=1-p$, 有

$$p^n+C_1^n p^{n-1}q+\cdots+C_k^n p^{n-k}q^k=\frac{\int_0^p x^{n-k-1}(1-x)^k\,dx}{\int_0^1 x^{n-k-1}(1-x)^k\,dx},\ \text{其中} C_i^n=\binom{n}{i}.$$

22. 设方程 $(x-\alpha)(x-\beta)=x^2+px+q=0$ 的两个根几何型地出现于 $-1\leqslant\alpha\leqslant 1$, $-1\leqslant\beta\leqslant 1$, 试求系数 p,q 的分布密度 $f_p(x)$ 及 $f_q(x)$.

23. 设 $p(\xi=k)=\frac{a^k}{(1+a)^{k+1}}$, $a>0$ 常数， $k=0,1,2,\cdots$. 试求 $E\xi,D\xi$.

24. 设 ξ 的分布密度是 $f(x)=cxe^{-h^2x^2}$, $x\geqslant 0$. 试求常数 c 及 $E\xi,D\xi$(h 为已知常数).

25. 设 ξ (i) 分布在 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 上，并具有密度 $f(x)=\frac{2}{\pi}\cos^2 x$;

(ii) 分布在 $(-1,1)$ 上，并具有密度 $f(x)=c(1-x^2)^\alpha$, $(\alpha>0)$.

试求： $E\xi$, $D\xi$.

26. 试证：对随机变数 ξ,η, $E\xi\eta = E\xi \cdot E\eta$ 或 $D(\xi+\eta) = D(\xi) + D(\eta)$ 成立的充要条件是 $r_{\xi\eta} = 0$ ($r_{\xi\eta}$ 是 ξ 与 η 的相关系数).

27. 设 $\xi = (\xi_1, \cdots, \xi_k)$ 具有多项分布

$$p(\xi_1 = x_1, \cdots, \xi_k = x_k) = \frac{n!}{x_1! \cdots x_{k+1}!} p_1^{x_1} \cdots p_{k+1}^{x_{k+1}},$$

$$x_{k+1} = n - x_1 - \cdots - x_k,$$

$$p_{k+1} = 1 - p_1 - \cdots - p_k.$$

试证： $E\xi_i = np_i$, $D\xi_i = np_i(1-p_i)$, $b_{ij} = -np_ip_j$ $(i \neq j)$.

28. 设 ξ 的数学期望是 μ, 方差是 σ^2. 试证：对任何正常数 λ, 有： $P(|\xi - \mu| \geqslant \lambda\sigma) \leqslant \frac{1}{\lambda^2}$.

29. 若中心矩 $c_{2r}, c_{2r+1}, c_{2r+2}$ 存在，试证 $(c_{2r+1})^2 \leqslant c_{2r}c_{2r+2}$.

30. 设 ξ,η 有二维正态分布， $E\xi = a_1$, $D\xi = 1$, $E\eta = a_2$, $D\eta = 1$. 试证 ξ,η 的相关系数 $r = \cos q\pi$, 其中

$$q = P((\xi - a_1)(\eta - a_2) < 0).$$

31. 设随机向量 (ξ,η) 服从二维正态分布 $E\xi = E\eta = 0$, $D\xi = D\eta = 1$, $r_{\xi\eta} = R$. 试证 $E\max(\xi,\eta) = \sqrt{\frac{1-R}{\pi}}$.

32. 设 X,Y 独立有相同分布 $N(0,\sigma)$, 令 $\xi = \alpha X + \beta Y$. $\eta = \alpha X - \beta Y$. 求 ξ 与 η 的相关系数 $r_{\xi\eta}$.

33. 设 X_1, X_2 独立有相同分布 $N(a,\sigma)$. 试证

$$E\max(X_1, X_2) = a + \frac{\sigma}{\sqrt{\pi}}.$$

34. 求证：对于随机变数 X 及正值上升函数 $f(x)$ $(x \geqslant 0)$, 若 $Ef(|X|) < \infty$, 则

$$P(|X| \geqslant x) \leqslant \frac{E(f(|X|))}{f(x)} \quad (x > 0).$$

35. 求证：若 $E(e^{aX}) < \infty$, $(a > 0)$, 则 $P(X \geqslant x) \leqslant e^{-ax}E(e^{aX})$.

36. 设 ξ 的分布函数为 $F(x) = \int_{-\infty}^{x} f(y)\,dy$, 求

(i) $\eta = e^{-\xi}$; (设 $\xi > 0$)

(ii) $\eta = \operatorname{tg}\xi$;

(iii) $\eta = \operatorname{arctg}\xi$

的分布函数 $F_\eta(x)$ 及密度 $f_\eta(x)$.

37. 求证：对取值于区间 (a,b) 内的随机变数 ξ, 恒成立不等式 $a \leqslant E\xi \leqslant b$, $D\xi \leqslant \left(\frac{b-a}{2}\right)^2$.

38. 设 ξ 只取非负整数值， $P(\xi = k) = p_k$, $k = 0, 1, 2, \cdots$. 定义母函数 $P(s) = \sum\limits_{k=0}^{\infty} p_k s^k$. 试证：

$$E\xi = P'(1), \ D\xi = P''(1) + P'(1) - [P'(1)]^2,$$

其中 $P'(1)$, $P''(1)$ 为 $P(s)$ 的一级，二级导数在 $s = 1$ 的值.

39. 对二项分布及 Poisson 分布分别求出 $P(s)$.

40. 设 ξ 的分布函数是 $F(x)$, 定义矩母函数 $m(t) = \int_{-\infty}^{\infty} e^{tx}\,dF(x) = E(e^{t\xi})$. 试证：如果 $m(t)$ 存在，则 $E\xi = m'(0)$, $E\xi^2 = m''(0)$.

41. 设 $\xi_1, \cdots, \xi_k$ 是独立同分布随机变数，期望为 0, 方差为 1. 试证：对 $\lambda > 0$, $P\left(\sum\limits_{i=1}^{k} \xi_i^2 \geqslant \lambda_k\right) \leqslant \frac{1}{\lambda}$.

42. 试证：如果密度函数对称，即如 $f(x)=f(-x)$, 又设奇数阶的半不变量存在，则它们一定是 0.

43. 设分布函数 $F(x)=\begin{cases}0, & x<a,\\ p, & a\leqslant x<b,\\ 1, & x\geqslant b.\end{cases}$ 求特征函数$f(t)$,

44. 对下列分布函数求特征函数： $F(x)=\begin{cases}0, & x<-a,\\ \dfrac{x+a}{2a}, & -a\leqslant x<a,\\ 1, & x\geqslant a.\end{cases}$

45. 对下列分布密度求特征函数： $g(x)=\frac{a}{2}e^{-a|x|}$, $a>0$.

46. 设 ξ 服从分布： $P(\xi=k)=q^kp$, $0<p<1$, $q=1-p$, $k=0,1,2,\cdots$. 求 ξ 的特征函数，并求 $E\xi$, $D\xi$.

47. 若特征函数是 $\frac{1}{1+t^2}$, 试证对应的分布密度是 $\frac{1}{2}e^{-|x|}$, $-\infty<x<\infty$.

48. 若特征函数是 $\frac{e^{it}(1-e^{nit})}{n(1-e^{it})}$, 试证对应的随机变数的密度矩阵为 $\begin{pmatrix}1 & 2 & \cdots & n\\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n}\end{pmatrix}$.

49. 对下列特征函数，求分布密度或分布函数

(i) $\frac{1}{1-it}$, (ii) $\frac{1}{(1-it)^n}$.

50. 试证：如特征函数满足 $f(t)=1+o(t^2)$, $t\to 0$, 则 $f(t)\equiv 1$.

51. 试证：特征函数 $f(t)$ 是实值的充要条件是 $f(t)$ 相应的分布函数 $F(x)$ 是对称的 (即 $F(x)$ 满足 $F(x)=1-F(-x-0)$).

52. 若 $f(t)$ 是分布函数 $F(x)$ 的特征函数，则对所有的 T 和 h 有

$$\frac{1}{h}\int_{T+h}^{T+2h}F(x)\,dx-\frac{1}{h}\int_{T}^{T+h}F(x)\,dx=\frac{1}{2\pi h}\int_{-\infty}^{\infty}\left(\frac{1-e^{-iTh}}{it}\right)^2e^{-itT}f(t)\,dt.$$

53. 如 ξ 有分布密度，则它特征函数当 $t\to\infty$ 时趋于 0.

54. 求证：对任何实值特征函数 $f(t)$, 以下二不等式成立：

$$1-f(2t)\leqslant 4(1-f(t)),$$
$$1+f(2t)\geqslant 2(f(t))^2.$$

55. 设 $\xi_1,\cdots,\xi_n$ 独立，有相同的几何分布 (见题 46). 试求 $\sum\limits_{i=1}^{n}\xi_i$ 的分布.

56. 设 $F(x)$ 是分布函数，有特征函数为 $f(t)$. 试证 $G(x)=\frac{1}{2h}\int_{x-h}^{x+h}F(y)\,dy$ 也是分布函数，而且特征函数为 $g(t)=\frac{\sin ht}{ht}f(t)$, 其中 $h>0$ 为常数.

57. 试证满足下列各等式的连续函数 $f(t)$ 是特征函数：

$$f(t)=f(-t),\ f(t+2a)=f(t),\ f(t)=\frac{a-t}{a},\ 0\leqslant t\leqslant a.$$

58. 对事件 A_i 定义随机变数 $\chi_{A_i}(\omega)=0$ 或 1, 视 $\omega\overline{\in}A$ 或 $\omega\in A$ 而定，试证事件 $A_1,\cdots,A_n$ 独立的充要条件是 $\chi_{A_1}(\omega),\cdots,\chi_{A_n}(\omega)$ 独立.

59. 巴士加 (Pascal) 分布：设贝努里试验进行到第 r 次成功出现为止 (每次试验中成功的概率为 p, $q=1-p$). 令 X 为试验进行的次数，则事件 $X=k$ 等价于 “第 k 次试验时出现成功，且在其前 $k-1$ 次试验中成功 $r-1$ 次”. 故

$$P(X=k)=p\binom{k-1}{r-1}p^{r-1}q^{k-r}=\binom{k-1}{r-1}p^rq^{k-r},\quad k=r,r+1,\cdots.$$

此分布叫巴士加分布. 当 $r=1$ 时，化为几何分布. 令 ξ_1 为第一次出现成功的试验次数， ξ_2 为第一次成功后到第二次成功止之间的试验次数， $\cdots$, 则 $\xi_1,\cdots,\xi_r$ 独立，有相同的参数为 p 的几何分布，后者的数学期望为 $1/p$, 方差为 q/p^2, 故由 $X=\xi_1+\cdots+\xi_r$, 得 $E(X)=r/p$, $D(X)=rq/p^2$.

60. 设 ξ 有指数分布，分布函数为 $F(x)=1-e^{-\alpha x}$,$(x\geqslant 0)$, $\alpha>0$. 则

$$P(\xi>s+t|\xi>s)=P(\xi\geqslant t)$$

这叫做无记忆性. 在可靠性问题中，把 ξ 理解为某元件的寿命，则上式表示：在已知寿命超过 s 的条件下，元件再继续工作 t 时以上的条件概率，等于它工作 t 时以上的无条件概率，而不依赖于它已经工作了 s 时以上. 上式极易证明，因为其左方等于 $e^{-\alpha(s+t)}/e^{-\alpha s}=e^{-\alpha t}$. 进一步，还可证明：指数分布是唯一的具有无记忆性的连续型分布；而几何分布则是唯一的具有无记忆性的集中在正整数上的离散型分布 (见 Feller 书 [15], 卷 1,340 页；卷二， 8 页). 指数分布所以广泛出现在马尔科夫过程，排队论与可靠性等问题中，无记忆性是重要原因之一.

61. 设 (ξ_1,ξ_2) 有二维正态分布，密度见 §2.7(9). 试证在 $\xi_1=x$ 的条件下， ξ_2 的条件密度为

$$f(y|x)=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{\frac{-1}{2\sigma_2^2(1-r^2)}\left[y-b-\frac{r\sigma_2}{\sigma_1}(x-a)\right]^2\right\}.$$

这是期望为 $b+(r\sigma_2/\sigma_1)(x-a)$, 方差为 $\sigma_2^2(1-r^2)$ 的一维正态分布的密度，类似求 $f(x|y)$.

62. 设 f_1 与 f_2 都是二维正态分布密度，期望皆为 0, 方差皆为 1, 但有不同的相关系数. 显然 $\frac{1}{2}(f_1+f_2)$ 也是二维概率分布密度，而且两个边沿分布为 $N(0,1)$ 正态的，但本身却不是二维正态的.

第三章 独立随机变数序列的极限定理

§3.1 四种收敛性

(一) 几乎处处收敛. 在数学分析和实变函数论中，我们已经看到，极限理论中 "收敛性" 概念极为重要，如果收敛的定义改变，那么全部极限理论都要相应地改变，概率论中也是这样. 因此，在叙述极限定理以前，必须先把收敛的定义弄清楚. 本书中用到的随机变数的收敛性共有四种 [1]：几乎处处收敛，依概率收敛，依分布收敛，$r-$ 阶收敛.

先作些准备，设 $(\Omega, \mathcal{F}, P)$ 为概率空间，对任意一列事件 $A_n, n=1,2,\cdots$, 回忆

$$\bigcup_{n=1}^{\infty} A_n = (\omega: \text{至少存在一整数 } n \geqslant 1, \text{使 } \omega \in A_n), \tag{1}$$

$$\bigcap_{n=1}^{\infty} A_n = (\omega: \omega \in A_n, \text{一切} n = 1,2,\cdots); \tag{2}$$

由此可证

$$\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = (\omega: \omega \text{ 属于无穷多个} A_n), \tag{3}$$

$$\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n = (\omega: \omega \text{ 属于几乎一切} A_n), \tag{4}$$

这里 "ω 属于几乎一切 A_n" 的详细内容是："存在一正整数 N, 使 $\omega \in A_n$, 一切 $n \geqslant N$" (N 一般依赖于 ω). 实际上，"$\omega \in \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$" 是说："对任何一个正整数 k, 必定存在一正整数 $n_k \geqslant k$, 使 $\omega \in A_{n_k}$", 于是 ω 属于一切 A_{n_k}, $k=1,2,\cdots$, 因而 ω 属于无穷多个 A_n, 故得证：

$$\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \subset (\omega: \omega \text{ 属于无穷多个} A_n). \tag{5}$$

把上面的推理反向而行，即知

$$\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \supset (\omega: \omega \text{ 属于无穷多个} A_n). \tag{6}$$

由 (5)(6) 即得证 (3).(4) 的证明类似.

作为可列多个事件 A_n 的和或交，$\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$ 及 $\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$ 也是事件.

---

[1] 对于学过实变函数论的读者，这四种收敛中本质上新的只是依分布收敛，故可扼要阅读本节内容，本章的主要任务是叙述大数定理与中心极限定理，§3.1 与 §3.2 是作准备工作.

设 $\xi_n(\omega), n=1,2,\cdots$, 及 $\xi(\omega)$ 都是定义在 $(\Omega,\mathcal{F},P)$ 上的随机变数. 当 $\omega\in\Omega$ 固定时, 我们采用数学分析中常数序列极限的定义来定义

$$\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega). \tag{7}$$

如果

$$P(\omega:\ \lim_{n\to\infty}\xi_n(\omega)=\xi(\omega))=1, \tag{8}$$

就说 $\{\xi_n(\omega)\}$ 几乎处处 (或概率 1) 收敛到 $\xi(\omega)$, 并记为

$$\lim_{n\to\infty}\xi_n=\xi,\ (\text{a.s.}) \tag{9}$$

这种收敛在 §2.10 中已初步介绍过.

注意 (8) 中括号里的 ω- 集是一事件, 因而 (8) 的左方是有意义的, 实际上

$$(\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega))=\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)\in\mathcal{F}. \tag{10}$$

(10) 中的等式确实成立, 因为 $\omega\in(\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega))$ 的充要条件是: 对任一正整数 m, 存在一正整数 N, 使 $|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}$, 对一切 $n\geqslant N$ 成立; 换句话说, 也就是: 对任一正整数 m, ω 属于几乎一切 A_n, 这里 $A_n=\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)$. 根据 (4), 这也正是 ω 属于 (10) 中等号右方集的充要条件.

引理 1 (波雷耳 - 肯塔里 (Borel-Gantelli) 引理) 设 $\{A_n\}$ 为事件列, $\sum\limits_{n=1}^{\infty}P(A_n)<\infty$, 则

$$P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right)=0. \tag{11}$$

证 令 $B_k=\bigcup\limits_{n=k}^{\infty}A_n$, $B_k\supset B_{k+1}$, 故

$$P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right)=\lim_{k\to\infty}P(B_k)\leqslant\lim_{k\to\infty}\sum_{n=k}^{\infty}P(A_n)=0.\quad\square \tag{12}$$

根据 (3), 引理 1 可改述为: 如 $\sum\limits_{n=1}^{\infty}P(A_n)<\infty$, 则属于无穷多个 A_n 的 ω 所成的集有概率为 0; 或者说, 只属于有穷多个 A_n 的 ω 所成的集的概率为 1. 通常把引理 1 直观地改述为: 如 $\sum\limits_{n=1}^{\infty}P(A_n)<\infty$, 则以概率 1 只出现有穷多个事件 A_n.

现在介绍一种补集运算: 设集 A 可表为一列集 A_n 经 "和" 与 "交" 即 "$\cup$" 与 "$\cap$" 运算得来, 那么 $\overline{A}$ 可由 $\overline{A}_n$ 经 "$\cap$" 与 "$\cup$" 得到. 只要把原来的 A_n 与 $\overline{A}_n$ 对换, 把 $\cup$ 与 $\cap$ 对换, 把 ϕ 与 Ω 对换. 这叫对偶原则. 例如, 设 $A=\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}A_n$, 则 $\overline{A}=\bigcup\limits_{k=1}^{\infty}\bigcap\limits_{n=k}^{\infty}\overline{A}_n$; 注意 $\overline{A}=(\omega:\ \omega$ 只属于有穷多个 $A_n)$.

(二) **依概率收敛**. 说随机变数列 $\{\xi_n(\omega)\}$ 依概率收敛于随机变数 $\xi(\omega)$, 如果对任意 $\varepsilon>0$, 有

$$\lim_{n\to\infty}P(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)=0 \tag{13}$$

记此收敛为

$$\xi_n \longrightarrow \xi,\ (\mathrm{P}) \tag{14}$$

显然，(13) 等价于

$$\lim_{n\to\infty} P(|\xi_n(\omega)-\xi(\omega)|<\varepsilon)=1. \tag{15}$$

定理 1 (i) $\xi_n \longrightarrow \xi$, (a.s.) 的充要条件是：对任意 $\varepsilon>0$, 有

$$\lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty}(|\xi_k-\xi|\geqslant\varepsilon)\right)=0; \tag{16}$$

(ii) 如 $\xi_n\to\xi$, (a.s.), 则 $\xi_n\to\xi$, (P).

证 (ii) 显然由 (i) 及 (13) 推出，只要证 (i).

必要：如在定点 ω 上有 $\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega)$, 则对任意 $\varepsilon>0$, 不等式 $|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon$ 不能对无穷多个 n 成立.

令 $A_n=\left(\omega:\ \bigcup\limits_{k=n}^{\infty}(|\xi_k(\omega)-\xi(\omega)|\geqslant\varepsilon\right)$, 则 $A_n\supset A_{n+1}$, 故由 §1.3 定理 2 及 $\xi_n\to\xi$, (a.s.) 得

$$\lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty}(|\xi_k-\xi|\geqslant\varepsilon)\right)=P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}(|\xi_k-\xi|\geqslant\varepsilon)\right)=0. \tag{17}$$

充分：由 (16) 及上式第一等号得

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\left(|\xi_k-\xi|\geqslant\frac{1}{m}\right)\right)=0. \tag{18}$$

注意对可列多个概率为 0 的事件 A_n 的和 $A=\bigcup\limits_{n=1}^{\infty}A_n$, 有 $P(A)\leqslant\sum\limits_{n=1}^{\infty}P(A_n)=0$, 即 $P(A)=0$, 故

$$P\left(\bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\left(|\xi_k-\xi|\geqslant\frac{1}{m}\right)\right)=0.$$

由对偶原则，即得

$$P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left(|\xi_k-\xi|<\frac{1}{m}\right)\right)=1.$$

由此及 (10) 推出 $\xi_n\to\xi$, (a.s.). □

(ii) 之逆不真：

例 1 取 $\Omega=(0,1]$, $\mathcal{F}$ 为 $(0,1]$ 中全体波雷耳子集所成 σ 代数，P 为勒贝格测度. 令

$$\eta_{11}(\omega)=1;\quad \eta_{21}(\omega)=\begin{cases}1, & \omega\in\left(0,\dfrac{1}{2}\right],\\ 0, & \omega\in\left(\dfrac{1}{2},1\right];\end{cases}\quad \eta_{22}(\omega)=\begin{cases}0, & \omega\in\left(0,\dfrac{1}{2}\right],\\ 1, & \omega\in\left(\dfrac{1}{2},1\right].\end{cases}$$

一般地，将 $(0,1]$ 分成 k 个等长区间，而令

$$\eta_{ki}(\omega)=\begin{cases}1, & \omega\in\left(\dfrac{i-1}{k},\dfrac{i}{k}\right]\\ 0, & \omega\overline{\in}\left(\dfrac{i-1}{k},\dfrac{i}{k}\right]\end{cases}\quad (i=1,2,\cdots,k;\ k=1,2,\cdots).$$

定义

$$\xi_1(\omega)=\eta_{11}(\omega),\quad \xi_2(\omega)=\eta_{21}(\omega),\quad \xi_3(\omega)=\eta_{22}(\omega),$$
$$\xi_4(\omega)=\eta_{31}(\omega),\quad \xi_5(\omega)=\eta_{32}(\omega),\qquad \cdots$$

$\{\xi_n(\omega)\}$ 是一列随机变数. 对任意 $\varepsilon>0$, 由于

$$P(|\eta_{ki}(\omega)|\geqslant\varepsilon)\leqslant\frac{1}{k},\tag{19}$$

故 $P(|\xi_n(\omega)|\geqslant\varepsilon)\to 0(n\to\infty)$, 即 $\xi_n\to 0$, (P); 然而对任一固定 $\omega\in\Omega$, 任一正整数 k, 恰有一 i, 使 $\eta_{ki}(\omega)=1$, 而对其余的 j 有 $\eta_{kj}(\omega)=0$. 由此知 $\{\xi_n(\omega)\}$ 中有无穷多个 1 及无穷多个 0, 于是 $\{\xi_n(\omega)\}$ 对每一 $\omega\in\Omega$ 都不收敛.

(三) **弱收敛**. 称分布函数列 $\{F_n(x)\}$ 为弱收敛的 [1], 如果存在一单调不减函数 $F(x)$, 使在 $F(x)$ 的每一连续点 x 上有:

$$\lim_{n\to\infty}F_n(x)=F(x).\tag{20}$$

这时也说 $\{F_n(x)\}$ 弱收敛到 $F(x)$, 并记为 $F_n(x)\to F(x)$, (w),

尽管 $\{F_n(x)\}$ 是分布函数列, 然而, 由下例看出, $F(x)$ 却未必是分布函数, 因此 "$\{F_n(x)\}$ 弱收敛到 **分布** 函数 $F(x)$" 和 "$\{F_n(x)\}$ 弱收敛到 $F(x)$" 两句话有重大的差别, 前者不仅要求后者成立, 而且还要求 $F(x)$ 是分布函数.

例 2 取

$$F_n(x)=\begin{cases}0, & 如 x<n,\\ 1, & 如 x\geqslant n;\end{cases}\qquad F(x)\equiv 0.\tag{21}$$

显然, $F_n(x)\to F(x)$, (w), 但 $F(x)$ 不是分布函数.

设随机变数 $\xi_n(\omega),\xi(\omega)$ 的分布函数分别为 $F_n(x),F(x)$, 如果 $F_n(x)\to F(x)$, (w), 就称 $\{\xi_n(\omega)\}$ 依分布收敛到 $\xi(\omega)$, 并记为 $\xi_n\to\xi$,(w).

注意, 我们只要求在 $F(x)$ 的连续点上 (20) 成立, 如果希望 (20) 对一切 $x\in R$ 成立, 那么要求就会过于苛刻, 结果连处处收敛的随机变数列也不能满足这一苛刻条件.

例 3 任意取一列常数 $\{c_n\}$, 使 $c_1>c_2>\cdots$,

$$\lim_{n\to\infty}c_n=c\ \ (c>-\infty).$$

令 $\xi_n(\omega)=c_n,\xi(\omega)=c$ (一切 ω). 显然, 对每一 ω 有 $\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega)$. 其次, $\xi_n(\omega)$ 及 $\xi(\omega)$ 的分布函数分别为

$$F_n(x)=\begin{cases}0, & x<c_n,\\ 1, & x\geqslant c_n;\end{cases}\qquad F(x)=\begin{cases}0, & x<c,\\ 1, & x\geqslant c.\end{cases}$$

$F_n(x)\to F(x)$, (w); 但在 $F(x)$ 的不连续点 c 上,

$$F_n(c)=0,\qquad F(c)=1.$$

[1] 弱收敛分布函数列 $\{F_n(x)\}$ 的极限函数 $F(x)$ 不唯一, 但右连续的极限函数唯一. 参看 §3.2 例 2 后的说明.

故 $\lim\limits_{n\to\infty} F_n(c) \neq F(c)$.

定理 2 设 $\xi_n \to \xi,(\mathrm{P})$, 则 $\xi_n \to \xi,(\mathrm{w})$.

证 对任意 $x \in R_1, y \in R_1$, 有

$$\begin{aligned}(\xi \leqslant \eta) &= (\xi_n \leqslant x, \xi \leqslant y) \cup (\xi_n > x, \xi \leqslant y)\\ &\subset (\xi_n \leqslant x) \cup (\xi_n > x, \xi \leqslant y)\\ F(y) &\leqslant F_n(x) + P(\xi_n > x, \xi \leqslant y),\end{aligned}$$

由于 $\xi_n \to \xi$, (P), 故对 $y < x$ 得

$$P(\xi_n > x, \xi \leqslant y) \leqslant P(|\xi_n - \xi| \geqslant x - y) \to 0, \quad (n \to \infty).$$

因此

$$F(y) \leqslant \varliminf_{n\to\infty} F_n(x);$$

类似可证：对 $x < z$, 有

$$\varlimsup_{n\to\infty} F_n(x) \leqslant F(z),$$

于是对 $y < x < z$,

$$F(y) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(z).$$

如果 x 是 $F(x)$ 的连续点，令 $y \to x, z \to x$, 得

$$F(x) = \lim_{n\to\infty} F_n(x). \quad \square$$

定理 2 的逆不真.

例 4 设 $\xi_1, \xi_2, \cdots$ 及 ξ 是相互独立的随机变量，有公共的密度矩阵为 $\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, 因而有相同的分布函数，故 $\xi_n \to \xi(\mathrm{w})$, 但对 $1 > \varepsilon > 0$

$$\begin{aligned}P(|\xi_n - \xi| > \varepsilon) &= P(\xi_n = 1, \xi = 0) + P(\xi_n = 0, \xi = 1)\\ &= P(\xi_n = 1)P(\xi = 0) + P(\xi_n = 0)P(\xi = 1)\\ &= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2},\end{aligned}$$

故 $\{\xi_n\}$ 不依概率收敛于 ξ.

定理 3 设分布函数列 $\{F_n(x)\}$ 弱收敛于连续的分布函数 $F(x)$, 则这收敛对 $x (\in R_1)$ 是均匀的.

证 由于 $F(x)$ 有界不下降， $F(-\infty) = 0$, $F(\infty) = 1$, 故对任意 $\varepsilon > 0$, 可找到 M, 使当 $x \geqslant M$ 时有

$$1 - F(x) < \varepsilon, \tag{22}$$

且当 $x \leqslant -M$ 时有

$$F(x) < \varepsilon. \tag{23}$$

由于 $F_n(x)$ 处处收敛于 $F(x)$, 故存在一正整数 N_1, 使当 $n > N_1$ 时，一方面有

$$|F_n(-M) - F(-M)| < \varepsilon,$$

由 (23) 得

$$F_n(-M) < 2\varepsilon; \tag{24}$$

另方面又有

$$|F_n(M) - F(M)| < \varepsilon,$$

由 (22) 得

$$1 - F_n(M) < 2\varepsilon \tag{25}$$

因此，对 $x < -M$, 如 $n \geqslant N_1$, 则由 (23),(24) 有

$$|F_n(x) - F(x)| \leqslant |F_n(x)| + |F(x)| \leqslant F_n(-M) + F(-M) < 3\varepsilon; \tag{26}$$

同样，对 $x > M$, 如 $n > N_1$, 则由 (22),(25) 有

$$\begin{aligned} |F_n(x) - F(x)| &= |(1 - F_n(x)) - (1 - F(x))| \\ &\leqslant |1 - F_n(x)| + |1 - F(x)| \leqslant 1 - F_n(M) + 1 - F(M) < 3\varepsilon. \end{aligned} \tag{27}$$

在有限闭区间 $[-M, M]$ 上， $F(x)$ 连续，故也均匀连续，因而在 $[-M, M]$ 上可找到 l 个点 $x_1, x_2, \cdots, x_l, x_1 = -M, x_l = M$ 使

$$F(x_{i+1}) - F(x_i) < \varepsilon, \ (i = 1, 2, \cdots, l-1) \tag{28}$$

还可找到 $N_2 > N_1$, 使在此 l 个点中的每一点上，当 $n > N_2$ 时有

$$|F_n(x_i) - F(x_i)| < \varepsilon. \tag{29}$$

于 $[-M, M]$ 中任取一 x, 则此 x 必属于某一 $[x_i, x_{i+1}]$. 因之当 $n > N_2$ 时，由 (29) 有

$$F_n(x) \leqslant F_n(x_{i+1}) < F(x_{i+1}) + \varepsilon, \tag{30}$$

及

$$F_n(x) \geqslant F_n(x_i) > F(x_i) - \varepsilon, \tag{31}$$

由此及 (30),(28) 得

$$F_n(x) - F(x) < F(x_{i+1}) - F(x) + \varepsilon \leqslant F(x_{i+1}) - F(x_i) + \varepsilon < 2\varepsilon; \tag{32}$$

同样由 (31),(28) 得

$$F_n(x) - F(x) > F(x_i) - F(x) - \varepsilon \geqslant F(x_i) - F(x_{i+1}) - \varepsilon > -2\varepsilon. \tag{33}$$

故当 $N > N_2$ 时由 (26),(27),(32),(33), 得对任意 $x \in R_1$, 有

$$|F_n(x) - F(x)| < 3\varepsilon. \quad \square$$

（四）r- **阶收敛**．设对随机变数 ξ_n 及 ξ 有 $E|\xi_n|^r<\infty$, $E|\xi|^r<\infty$, 其中 $r>0$ 为常数，如果

$$\lim_{n\to\infty} E|\xi_n-\xi|^r=0, \tag{34}$$

则称 $\{\xi_n\}$ r- 阶收敛于 ξ, 记为 $\xi_n\to\xi(\mathrm{r})$.

2- 阶收敛亦称均方收敛，是最重要的特殊情形．

定理 4 如 $\xi_n\to\xi,(\mathrm{r})$, 则 $\xi_n\to\xi$, (P).

证 由 §2.10 引理 1, 对 $\varepsilon>0$, 有

$$P(|\xi_n-\xi|\geqslant\varepsilon)\leqslant\frac{E|\xi_n-\xi|^r}{\varepsilon^r},$$

再利用 (34) 即可． □

定理 4 的逆不真．

例 5 $(\Omega,\mathcal{F},P)))$ 如例 1 所取，令

$$\xi_n(\omega)=\begin{cases} n^{\frac{1}{r}}, & \text{如 } \omega\in\left(0,\dfrac{1}{n}\right], \\ 0, & \text{如 } \omega\overline{\in}\left(0,\dfrac{1}{n}\right]; \end{cases}$$

$$\xi(\omega)=0,\qquad (\text{一切 } \omega).$$

显然， $\xi_n(\omega)\to\xi(\omega)$, 一切 ω, 故

$$\xi_n\to\xi,\ (\text{a.s.});\ \ \xi_n\to\xi,\ (\text{P}).$$

然而 $E|\xi_n-\xi|^r=n\cdot\frac{1}{n}=1$ 不趋于 0.

现在把四种收敛的关系总结如下：

几乎处处收敛 $\Longrightarrow$ 依概率收敛 $\Longrightarrow$ 依分布收敛．

r- 阶收敛 $\Longrightarrow$ 依概率收敛 $\Longrightarrow$ 依分布收敛．

这里 “$\Longrightarrow$” 表 “可推出”．至于几乎处处收敛与 r- 阶收敛之间则无蕴含关系，按前者收敛而按后者不收敛的例见例 5, 依后者收敛而不依前者收敛的例见例 1 (注意那里 $E|\eta_{ki}-0|^r=\frac{1}{k}\to 0, k\to\infty$) 最后，由 $\xi_n\to\xi$, (P), 虽不能推出 $\xi_n\to\xi$, (a.s.), 但必存在 $\{\xi_n\}$ 的子列 $\{\xi_{k_n}\}$, 使 $\xi_{k_n}\to\xi$, (a.s.), $(n\to\infty)$; 这结论以后不用，故证明从略．

§3.2 分布函数列与特征函数列

（一）**弱收敛的充要条件**．我们来研究分布函数列 $\{F_n(x)\}$ 弱收敛的充要条件，由于分布函数与特征函数一一对应，自然希望通过特征函数来表达这一条件．在讨论过程中，我们同时还引进了特征函数列的广义均匀收敛性．以后会看到，研究这两种函数列收敛性间的关系是很重要的．

引理 1 为了使分布函数列 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 只需在某一稠于 R_1 的可列集 D 上有

$$\lim_{n\to\infty} F_n(y)=F(y)\ \ (y\in D). \tag{1}$$

证 对任意 $x \in R_1$, 取 $y \in D$, $z \in D$, 使 $y < x < z$. 于是

$$F_n(y) \leqslant F_n(x) \leqslant F_n(z).$$

由 (1) 得

$$F(y) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(z).$$

如 x 是 $F(x)$ 的连续点，于上式中令 $y \to x-0$, $z \to x+0$, 得

$$F(x) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x).$$

故 $F(x) = \lim_{n\to\infty} F_n(x)$. □

定理 1 (海来 Helly 第一定理) 任一分布函数列 $\{F_n(x)\}$ 必定含一弱收敛于某函数 $F(x)$ 的子列，而且 $F(x)$ 单调不减，右连续. $0 \leqslant F(x) \leqslant 1$.

证 任取一稠于 R_1 的可列集 $D = \{y_n\}$, 数列 $\{F_n(y_1)\}$ 有界，故必含一收敛于某极限 $G(y_1)$ 的子列 $\{F_{1n}(y_1)\}$. 同样，由于数列 $\{F_{1n}(y_1)\}$ 有界，所以必含一收敛于某极限 $G(y_2)$ 的子列 $\{F_{2n}(y_2)\}$. 既然函数列 $\{F_{2n}(x)\}$ 是 $\{F_{1n}(x)\}$ 的子列，可见同时有

$$\lim_{n\to\infty} F_{2n}(y_1) = G(y_1), \quad \lim_{n\to\infty} F_{2n}(y_2) = G(y_2).$$

继续此手续 k 次而得子列 $\{F_{kn}(x)\}$, 对此子列，等式

$$\lim_{n\to\infty} F_{kn}(y_r) = G(y_r)$$

对一切 $r = 1, 2, \cdots, k$ 都成立.

现在取对角线序列 $\{F_{nn}(x)\}$, 由于它是 $\{F_{1n}(x)\}$ 的子列，故 $\lim_{n\to\infty} F_{nn}(y_1) = G(y_1)$; 又由于 $\{F_{nn}(x)\}(n \geqslant 2)$ 是 $\{F_{2n}(x)\}$ 的子列，所以 $\lim_{n\to\infty} F_{nn}(y_2) = G(y_2)$; 一般地，因为 $\{F_{nn}(x)\}(n \geqslant k)$ 是 $\{F_{kn}(x)\}$ 的子列，从而 $\lim_{n\to\infty} F_{nn}(y_k) = G(y_k)$. 总之，我们证明了： $\{F_n(x)\}$ 含一子列 $\{F_{nn}(x)\}$, 满足

$$\lim_{n\to\infty} F_{nn}(y) = G(y), \text{ 一切} y \in D, \tag{2}$$

根据 $F_{nn}(x)$ 的单调不减性及 $0 \leqslant F_{nn}(x) \leqslant 1$, 知 $G(y)$ 在 D 上也单调不减而且 $0 \leqslant G(y) \leqslant 1$.

现在对任一 $x \in R_1$, 令

$$\overline{G}(x) = \inf_{\binom{y \geqslant x}{y \in D}} G(y). \tag{3}$$

显然，定义在 R_1 上的函数 $\overline{G}(x)$ 单调不减， $0 \leqslant \overline{G}(x) \leqslant 1$, 而且

$$\overline{G}(x) = G(x) \quad (x \in D)$$

由引理 1 知

$$F_{nn}(x) \longrightarrow \overline{G}(x), \text{ (w)}$$

在 $\overline{G}(x)$ 的不连续点的集 A 上，修改 $\overline{G}(x)$ 的值使 $\overline{G}(x)$ 右连续；就是说，定义

$$F(x)=\begin{cases}\overline{G}(x), & x\overline{\in}A,\\ \lim\limits_{y\to x+0}\overline{G}(y), & x\in A.\end{cases}\tag{4}$$

显然 $F(x)$ 右连续，单调不减， $0\leqslant F(x)\leqslant 1$, 而且

$$F_{nn}(x)\longrightarrow F(x),\ (\mathrm{w}).\ \square$$

(二) 两个重要的收敛定理

定理 2 (正极限定理) 设分布函数列 $\{F_n(x)\}$ 弱收敛于某分布函数 $F(x)$, 则相应的特征函数列 $\{f_n(t)\}$ 收敛于相应的特征函数 $f(t)$, 而且这收敛在每一有限 t 区间内是均匀的.

证 对任意 $\varepsilon>0$, 取正数 a, 使 a 及 $-a$ 都是 $F(x)$ 的连续点，而且

$$F(-a)<\frac{\varepsilon}{10},\ \ 1-F(a)<\frac{\varepsilon}{10}.\tag{5}$$

次取正整数 N_1, 使当 $n\geqslant N_1$ 时有

$$|F_n(-a)-F(-a)|<\frac{\varepsilon}{10},\ \ |F_n(a)-F(a)|<\frac{\varepsilon}{10}.\tag{6}$$

于是对 $n\geqslant N_1$ 有

$$F_n(-a)<\frac{\varepsilon}{5},\ \ 1-F_n(a)<\frac{\varepsilon}{5}.\tag{7}$$

令 $A_1=(-\infty,-a]$, $A_2=(-a,a]$, $A_3=(a,\infty)$, 并记

$$\int_A e^{itx}(dF_n(x)-dF(x))=\int_A e^{itx}dF_n(x)-\int_A e^{itx}dF(x).$$

则

$$|f_n(t)-f(t)|\leqslant\sum_{j=1}^{3}\left|\int_{A_j}e^{itx}(dF_n(x)-dF(x))\right|.\tag{8}$$

由于 $|e^{itx}|=1$,

$$\left|\int_A e^{itx}(dF_n(x)-dF(x))\right|\leqslant\int_{A_j}dF_n(x)+\int_{A_j}dF(x)=F_n(-a)+F(-a).$$

当 $n\geqslant N_1$ 时，由 (5),(7) 得

$$\left|\int_A e^{itx}(dF_n(x)-dF(x))\right|\leqslant\frac{3\varepsilon}{10}.\tag{9}$$

同理

$$\left|\int_{A_3}e^{itx}(dF_n(x)-dF(x))\right|\leqslant\frac{3\varepsilon}{10}.\tag{10}$$

如能证对任意 $T>0$, 存在正整数 N_2, 使对一切 $t\in[-T,T]$, 有

$$\left|\int_{A_2}e^{itx}(dF_n(x)-dF(x))\right|<\frac{2\varepsilon}{5}.\tag{11}$$

则当 $n \geqslant N_1 + N_2$ 时，对一切 $t \in [-T, T]$, 由 (8),(9),(10),(11) 得

$$|f_n(t) - f(t)| < \varepsilon.$$

于是定理得证，故剩下来只要证 (11).

取点 x_k： $-a = x_0 < x_1 < \cdots < x_M = a$, 使每一 x_k 都是 $F(x)$ 的连续点，而且

$$l = \max_{1 \leqslant k \leqslant M}(x_k - x_{k-1}) < \frac{\varepsilon}{10T}, \quad T = 2a. \tag{12}$$

令 $B_k = (x_{k-1}, x_k]$, $k = 1, 2, \cdots, M$, 则

$$\begin{aligned}\left|\int_{A_2} e^{itx}(dF_n(x) - dF(x))\right| &\leqslant \left|\sum_{k=1}^{M}\int_{B_k} e^{itx_k}(dF_n(x) - dF(x))\right| \\ &+ \left|\sum_{k=1}^{M}\left[\int_{B_k}(e^{itx_k} - e^{itx})dF_n(x) - \int_{B_k}(e^{itx_k} - e^{itx})dF(x)\right]\right|.\end{aligned} \tag{13}$$

根据不等式

$$|e^{it(u+v)} - e^{itu}| = |e^{itv} - 1| = \left|it\int_0^v e^{its}\,ds\right| \leqslant |tv|. \ (v \geqslant 0) \tag{14}$$

可见对 $t \in [-T, T]$, 有

$$\sum_{k=1}^{M}\left|\int_{B_k}(e^{itx_k} - e^{itx})dF_n(x)\right| \leqslant \sum_{k=1}^{M} lT\int_{B_k} dF_n(x) < \frac{\varepsilon}{10}\sum_{k=1}^{M}\int_{B_k} dF_n(x) \leqslant \frac{\varepsilon}{10}.$$

把 $F_n(x)$ 换为 $F(x)$ 后上式也成立，故由 (13) 得

$$\left|\int_{A_2} e^{itx}(dF_n(x) - dF(x))\right| \leqslant \sum_{k=1}^{M}\left|\int_{B_k} e^{itx_k}(dF_n(x) - dF(x))\right| + \frac{\varepsilon}{5}. \tag{15}$$

然而

$$\begin{aligned}\left|\int_{B_k} e^{itx_k}(dF_n(x) - dF(x))\right| &= |e^{itx_k}[F_n(x_k) - F_n(x_{k-1}) - F(x_k) + F(x_{k-1})]| \\ &\leqslant |F_n(x_k) - F(x_k)| + |F_n(x_{k-1}) - F(x_{k-1})|.\end{aligned} \tag{16}$$

由于每 x_k 都是 $F(x)$ 的连续点，由假定，存在 N_2, 当 $n \geqslant N_2$ 时，有

$$|F_n(x_k) - F(x_k)| < \frac{\varepsilon}{10M} \qquad k = 0, 1, \cdots, M.$$

代入 (16) 得

$$\left|\int_{B_k} e^{itx_k}(dF_n(x) - dF(x))\right| < \frac{\varepsilon}{5M}.$$

以此式代入 (15), 即知对一切 $n \geqslant N_2$,(11) 式成立. □

定理 3 (逆极限定理) 设特征函数列 $\{f_n(t)\}$ 收敛于某连续函数 $f(t)$, 则相应的分布函数列 $\{F_n(x)\}$ 弱收敛于某一分布函数 $F(x)$, 而且 $f(t)$ 是 $F(x)$ 的特征函数.

证 由定理 1, 存在子列 $\{F_{n_k}(x)\}$ 弱收敛于某右连续, 单调不减函数 $F(x)$, $0 \leqslant F(x) \leqslant 1$, 试证此 $F(x)$ 是分布函数, 即还要证 $F(-\infty)=0$, $F(\infty)=1$. 显然,

$$0 \leqslant F(-\infty) \leqslant 1, \quad 0 \leqslant F(\infty) \leqslant 1.$$

故只要证 $F(\infty)-F(-\infty)=1$.

如说不然, 则

$$a = F(\infty) - F(-\infty) < 1. \tag{17}$$

由于 $f(0)=\lim\limits_{n\to\infty} f_n(0)=1$, 而且 $f(t)$ 连续, 故对满足 $0<\varepsilon<1-a$ 的 ε, 存在充分小的正数 τ, 使

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)\,dt\right| > 1-\frac{\varepsilon}{2} > a+\frac{\varepsilon}{2}. \tag{18}$$

既然 $F_{n_k}(x) \to F(x)$, (w), 由 (17) 推出, 可取 $b > \frac{4}{\varepsilon\tau}$, 使 $-b$ 及 b 都是 $F(x)$ 的连续点, 还可取正整数 K, 使对一切 $k \geqslant K$, 有

$$a_k = F_{n_k}(b) - F_{n_k}(-b) < a + \frac{\varepsilon}{4}.$$

由于 $f_{n_k}(t)$ 是 $F_{n_k}(x)$ 的特征函数, 故

$$\begin{aligned}\left|\int_{-\tau}^{\tau} f_{n_k}(t)\,dt\right| &= \left|\int_{-\infty}^{\infty}\left[\int_{-\tau}^{\tau} e^{itx}\,dt\right] dF_{n_k}(x)\right| \\ &\leqslant \left|\int\limits_{(-b,b]}\left[\int_{-\tau}^{\tau} e^{itx}\,dt\right] dF_{n_k}(x)\right| + \left|\int\limits_{R_1/(-b,b]}\left[\int_{-\tau}^{\tau} e^{itx}\,dt\right] dF_{n_k}(x)\right|\end{aligned} \tag{19}$$

考虑下列二不等式

$$\left|\int_{-\tau}^{\tau} e^{itx}\,dt\right| \leqslant \int_{-\tau}^{\tau} |e^{itx}|\,dt = 2\tau;$$

$$\left|\int_{-\tau}^{\tau} e^{itx}\,dt\right| = \left|\left[\frac{e^{itx}}{ix}\right]_{-\tau}^{\tau}\right| = \frac{2}{|x|}|\sin\tau x| \leqslant \frac{2}{|x|} \leqslant \frac{2}{b}\ (\text{如}|x| \geqslant b),$$

代入 (19), 即得

$$\left|\int_{-\tau}^{\tau} f_{n_k}(t)\,dt\right| \leqslant 2\tau a_k + \frac{2}{b}.$$

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f_{n_k}(t)\,dt\right| \leqslant a_k + \frac{1}{b\tau} \leqslant a_k + \frac{\varepsilon}{4} < a + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = a + \frac{\varepsilon}{2}.$$

令 $k\to\infty$, 由积分控制收敛定理得

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)\,dt\right| \leqslant a + \frac{\varepsilon}{2}.$$

此与 (18) 矛盾, 因而证明了 $F(x)$ 是一分布函数.

由定理 2 推知 $f(t)$ 是 $F(x)$ 的特征函数.

不仅子列 $F_{n_k}(x) \to F(x)$, (w), 其实 $\{F_n(x)\}$ 也弱收敛到同一分布函数 $F(x)$. 如说不然，必存在一子列 $\{F_{m_k}(x)\}$, 它弱收敛于某极限函数 $G(x)$, 后者至少在一连续点上不等于 $F(x)$. 但用上面同样的证法知 $G(x)$ 也是分布函数，而且也有特征函数为 $f(t)$, 根据唯一性定理 (§2.11 系 1), 知 $F(x) = G(x)$ 而导出矛盾. □

注 1. 定理 3 中 $f(t)$ 的连续性条件可放宽为 "$f(t)$ 在点 $t = 0$ 连续", 因为这仍可保证 (18) 成立，从而定理 3 的结论仍成立.

*(三) **弱收敛的各种等价条件**. 称特征函数列 $\{f_n(t)\}$ 为广义均匀收敛的，如果存在一个函数 $f(t)$, 使对每一 $t \in R_1$, 有

$$\lim_{n\to\infty} f_n(t) = f(t),$$

而且这收敛对每一有限区间 $[c, d]$ 中的 t 是均匀的 (即对任意 $\varepsilon > 0$, 任意有限区间 $[c, d]$, 存在正整数 $N = N(\varepsilon, c, d)$, 使对一切 $t \in [c, d]$, 当 $n \geqslant N$ 时，有 $|f_n(t) - f(t)| < \varepsilon$); 这时也说 $\{f_n(t)\}$ 广义均匀收敛到 $f(t)$.

注 2. 由于 $f_n(t)$ 连续，如 $\{f_n(t)\}$ 广义均匀收敛到 $f(t)$, 则 $f(t)$ 必定是连续函数.

系 1 设分布函数列 $\{F_n(x)\}$ 对应的特征函数列为 $\{f_n(t)\}$, 则下列四条件等价：

1° $\{F_n(x)\}$ 弱收敛于某分布函数 $F(x)$,

2° $\{f_n(t)\}$ 收敛到某函数 $f(t)$, $f(t)$ 在点 0 连续,

3° $\{f_n(t)\}$ 收敛到某连续函数 $f(t)$,

4° $\{f_n(t)\}$ 广义均匀收敛到某函数 $f(t)$.

当任一条件满足时，$f(t)$ 是 $F(x)$ 的特征函数.

证 由注 2, 自 4° 得 3°; 由 3° 显然得 2°; 由注 1, 自 2° 得 1°; 由定理 2, 自 1° 得 4°. 于是得证此四条件等价，最后一结论由定理 3 得到. □

由系 1 可见："分布函数列弱收敛于分布函数" 等价于 "特征函数列广义均匀收敛于特征函数".

由系 1 及 §2.11(i) 还可见，特征函数列的极限函数 $f(t)$ 在 $t = 0$ 的连续性，导致 $f(t)$ 在全 R_1 上的均匀连续性. 但 "$f(t)$ 在 $t = 0$ 的连续性" 却是不可少的条件.

*(四) **例**.

例 1 设

$$f_n(t) = \frac{\sin nt}{nt}, \quad (n = 1, 2, \cdots).$$

$\{f_n(t)\}$ 是一列特征函数 ($f_n(0) = 1$). 实际上,

$$\frac{\sin nt}{nt} = \frac{1}{2n}\int_{-n}^{n} e^{itx}\,dx = \int_{-\infty}^{\infty} e^{itx}\varphi_n(x)\,dx,$$

其中

$$\varphi_n(x) = \begin{cases} \dfrac{1}{2n}, & \text{如} x \in [-n, n], \\ 0, & \text{如} x \overline{\in} [-n, n] \end{cases} \tag{20}$$

是分布函数

$$F_n(x)=\begin{cases}0, & \text{如}x\leqslant -n,\\ \dfrac{x+n}{2n}, & \text{如}-n<x<n,\\ 1, & \text{如}x\geqslant n\end{cases} \tag{21}$$

的密度函数. 显然，对每 t, $\lim\limits_{n\to\infty} f_n(t)=f(t)$, 这里

$$f(t)=\begin{cases}1, & \text{如}t=0,\\ 0, & \text{如}t\neq 0.\end{cases}$$

$f(x)$ 在 0 点不连续，也不是特征函数.

附带指出，对 (21) 中 $F_n(x)$, 极限函数

$$F(x)=\lim_{n\to\infty} F_n(x)=\frac{1}{2}\quad(\text{一切}x\in R_1)$$

不是一分布函数 (见图 1), 读者可作出 $\varphi_n(x)$ 的图.

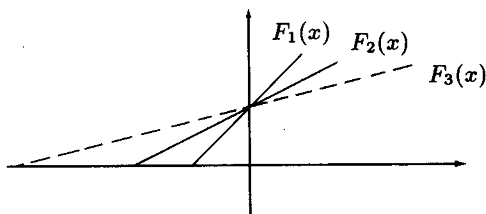


图 1

例 2 考虑正态分布函数

$$F_n(x)=\frac{n}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{n^2y^2}{2}}\,dy\quad(n=1,2,\cdots). \tag{22}$$

对应的特征函数列为

$$f_n(t)=e^{-\frac{t^2}{2n^2}}$$

显然， $F_n(x)\to F(x)$, (w)

$$F(x)=\begin{cases}0, & x<0,\\ 1, & x\geqslant 0.\end{cases}$$

$F(x)$ 是一分布函数，又

$$f_n(t)\to 1=\int_{-\infty}^{\infty} e^{ixt}\,dF(x).$$

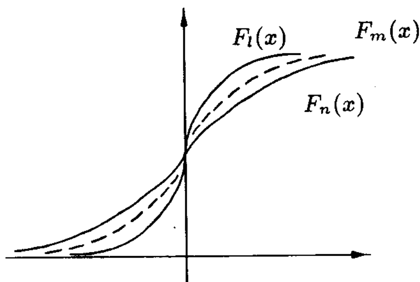


图 2 正态分布函数： $l>m>n$

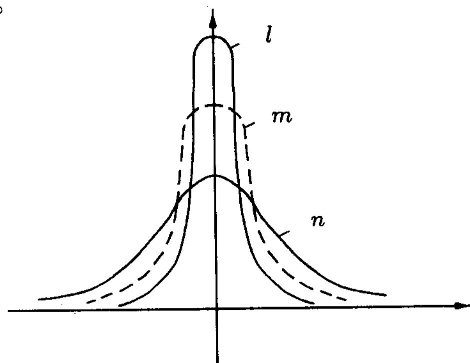


图 3 正态分布密度： $l>m>n$

对不同的 n 作出 $F_n(x)$ 及其密度的图 (见图 2 与图 3), 事情的本质就看得更清楚.

这个例子还说明, $\{F_n(x)\}$ 可弱收敛于二不相同的函数, 因为 (22) 中的 $\{F_n(x)\}$ 还弱收敛于 $G(x)$,

$$G(x)=\begin{cases}0, & x<0,\\ 1/2, & x=0\\ 1, & x>0.\end{cases}$$

不过, 如要求极限函数右连续, 那么极限函数必唯一, 这由下面的一般结果推出.

设 $\{F_n(x)\}$ 为任一列分布函数, 而且 $F_n(x)\to G_1(x)$, (w), $F_n(x)\to G_2(x)$, (w), 又 $G_1(x)$ 与 $G_2(x)$ 都右连续, 则 $G_1(x)=G_2(x)$, 一切 $x\in R_1$. 实际上, 由于 $G_1(x), G_2(x)$ 都单调不减, 值在 $[0,1]$ 之中, 故二者的不连续点所成的集 D 可数, 从而 $\overline{D}$ 处处稠密. 既然

$$G_1(x)=\lim_{n\to\infty}F_n(x)=G_2(x),\quad x\in\overline{D},$$

对任意 $y\in D$, 取 $x_n\in\overline{D}$, $x_n>y$, $x_n\to y$, 由右连续性, 立得

$$G_1(y)=\lim_{n\to\infty}G_1(x_n)=\lim_{n\to\infty}G_2(x_n)=G_2(y).$$

例 3 试根据已给的特征函数列 $\{f_j(t)\}$, $j=1,2,\cdots$, 利用定理 3 来造新的特征函数.

(i) 设 $f_j(t)$ 对应的分布函数为 $F_j(x)$, $\{a_j\}$ 为已给的常数列, $a_j\geqslant 0$, 显然, 函数

$$f(t)=\sum_{j=1}^{n}a_jf_j(t),\quad 其中\sum_{j=1}^{n}a_j=1. \tag{23}$$

也是一特征函数, 其实它是分布函数 $\sum\limits_{j=1}^{n}a_jF_j(x)$ 的特征函数. 由此可见, n 个特征函数的凸线性组合 (即组合系数 $a_j\geqslant 0$ 而且 $\sum\limits_{j=1}^{n}a_j=1$ 的线性组合) 是特征函数.

(ii) 今设非负项级数 $\sum\limits_{j=1}^{\infty}a_j=A, 0<A<\infty$, 于是当 n 充分大后, $\sum\limits_{j=1}^{n}a_j=A_n>0$. 由上述, 知

$$\varphi_n(t)=\sum_{j=1}^{n}\frac{a_j}{A_n}f_j(t)$$

是特征函数; 由于 $|f_j(t)|\leqslant 1$, $f_j(t)$ 连续, 故当 $n\to\infty$, 时, $\phi_n(t)$ 均匀收敛于某连续函数 $\varphi(t)$. 由定理 3 知 $\varphi(t)$ 也是特征函数, 而且它对应的分布函数为

$$\sum_{j=1}^{\infty}\frac{a_j}{A}F_j(x).$$

(iii) n 个特征函数的积 $\prod\limits_{j=1}^{n}f_j(t)$ 也是一特征函数. 实际上, 由 §2.7 定理 2, 可以造 n 个独立的随机变数 $\xi_j, j=1,\cdots,n$, 使 ξ_j 的分布函数为 $f_j(t)$ 所对应的分布函数 $F_j(x)$, 于是 $\sum\limits_{j=1}^{n}\xi_j$ 的特征函数是 $\prod\limits_{j=1}^{n}f_j(t)$. 特别, 知 $(f_1(t))^n$ 是 n 个独立同分布的随机变数的和的特征函数. 其次, 由于 $\overline{f_1(t)}=f_1(-t)$, 而由 §2.11 (iii), $f_1(-t)$ 是 $-\xi_1$ 的特征函数, 故

$$|f_1(t)|^2=f_1(t)f_1(-t) \tag{24}$$

是 $\eta_1-\xi_1$ 的特征函数，η_1 与 ξ_1 独立，有相同的分布函数 $F_1(x)$.

(iv) 取系数不全为 0 的幂级数 $g(x)=\sum\limits_{j=1}^{\infty}b_jx^j$，$b_j\geqslant 0$, 设它在 $x=1$ 收敛，于是

$$g(1)=\sum_{j=0}^{\infty}b_j>0.$$

任取一特征函数 $f(t)$, 由 (iii) 知 $(f(t))^j$ 是特征函数，由 (ii) 知

$$\varphi(t)=\frac{1}{g(1)}g(f(t))=\sum_{j=0}^{\infty}\frac{b_j}{g(1)}(f(t))^j \tag{25}$$

也是一特征函数 (注意，$(f(t))^\circ=1$ 是集中于 0 点的分布的特征函数).

例如，取 $f(t)=e^{it}$, 它对应的分布集中在点 1 上，又令 $g(x)=e^{\lambda x}$, $\lambda>0$, 由 (25) 知

$$\varphi(t)=\frac{1}{e^\lambda}e^{\lambda e^{it}}=e^{\lambda(e^{it}-1)}$$

是特征函数. 其实 $\varphi(t)$ 对应于参数为 λ 的普阿松分布 (§2.11, 例 2).

例 4 (随机端点问题) 长为 l 的直线的一端在原点 O, 与 x 轴的交角 α_1 是随机的. 然后以另一端为中心改换一次方向，新方向与 x 轴的交角为 α_2; 如此继续，共改变 n 次. 试求最后一次移动后的端点的横坐标 X_n 的分布 (下面绘出 X_7 的图，见图 4); 并研究当 $n\to\infty$ 时 X_n 的极限分布. 这里假定 $\{\alpha_j\}$ 独立，有相同的均匀分布密度

$$g(\alpha)=\begin{cases}0, & 如\alpha\overline{\in}[0,2\pi],\\ \dfrac{1}{2\pi}, & 如\alpha\in[0,2\pi].\end{cases}$$

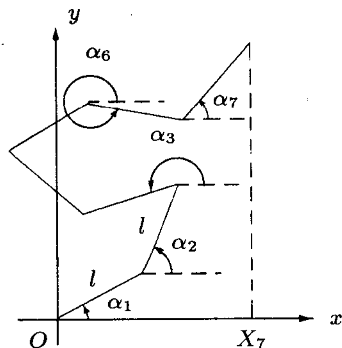


图 4

解

$$X_n=\sum_{j=1}^{n}l\cos\alpha_j.$$

因 $l\cos\alpha_j$ 的特征函数为

$$f(t)=\frac{1}{2\pi}\int_0^{2\pi}e^{itl\cos\alpha}d\alpha=J_0(tl),$$

其中 $J_0(x)$ 是第一类 0 级 Bessel 函数，它是偶函数. X_n 的特征函数为

$$f_n(t)=[J_0(tl)]^n.$$

由此得知 X_n 的密度为

$$p_n(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}[J_0(tl)]^n dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}[J_0(tl)]^n\cos(tx)\,dt.$$

今设 $n\to\infty$, 同时 $l\to 0$, 但二者有关系 $l=\frac{k}{\sqrt{n}}$, k 为正常数，$k=l\sqrt{n}$. 利用 $J_0(x)$ 的级数形

式，得

$$\begin{aligned} f_n(t) &= \left[1-\frac{(lt)^2}{2^2}+\frac{(lt)^4}{2^2\cdot 4^2}\cdots\right]^n \\ &= \left[1-\frac{k^2}{2^2\cdot n}t^2+\frac{k^4}{2^2\cdot 4^2\cdot n^2}t^4\cdots\right]^n \\ &\approx \left[1-\frac{k^2}{2^2\cdot n}t^2\right]^n \longrightarrow e^{-\frac{k^2t^2}{4}}\ (n\to\infty). \end{aligned}$$

故极限密度为

$$p(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-k^2t^2/4}\cos(tx)\,dt=\frac{1}{k\sqrt{\pi}}e^{-x^2/k^2}=\frac{1}{\lambda\sqrt{2\pi}}e^{-x^2/2\lambda^2},$$

这里 $\lambda=k/\sqrt{2}=l\sqrt{\frac{n}{2}}$. 因此， X_n 渐近地有 $N(0,l\sqrt{\frac{n}{2}})$ 正态分布.

这个例子在研究分子链等问题中有用. 参看 [46],33-39 页.

§3.3 大数定理与强大数定理

(一) 问题的一般提法. 我们知道，概率法则总是在对大量随机现象的考察中才能显现出来. 为了研究“大量”的随机现象，常常采用极限的形式，这就引导到极限定理的研究. 极限定理的内容很广泛，其中最重要的有两种：大数定理与中心极限定理. 我们从前者开始. 实际经验告诉我们，掷一个结构正常的硬币，虽然不能准确预言掷出的结果，但如独立地连掷 n 次，当 n 充分大后，出现正面的频率 $\frac{\eta_n}{n}$ 与 $\frac{1}{2}$ 很接近， η_n 表在此 n 次中出现正面的总次数. 这类事实可以如下直观地解释：要从随机现象中去寻求必然的法则，应该研究大量的随机现象，因为大量的随机现象里，各自的偶然性在一定程度上可以相互抵消，相互补偿，因而有可能显示必然的法则来.

于是现实向我们提出一个理论问题：在一般的贝努里实验中，每次实验时设 A 出现的概率为 $p, 0<p<1$, 以 η_n 表前 n 次实验里 A 出现的次数，能否从数学上严格证明：

$$\frac{\eta_n}{n}\longrightarrow p? \tag{1}$$

这问题也可改述如下：考虑独立随机变数列 $\{\xi_k\}$, 其中

$$\xi_k=\begin{cases}1, & \text{如 } A \text{ 在第 } k \text{ 次试验时出现},\\ 0, & \text{反之},\end{cases} \tag{2}$$

求证

$$\frac{\sum\limits_{k=1}^{n}\xi_k}{n}\longrightarrow p.\ (\text{注意}p=E\xi_k) \tag{3}$$

或者，更广泛些，我们提出下列一般问题：

设已给随机变数列 $\{\xi_k\}$ 及二列常数 $\{C_k\},\{D_k\},D_k\neq 0,k=1,2,\cdots$. 试研究随机变数列 $\{\zeta_n\}$ 的收敛性，这里

$$\zeta_n=\frac{1}{D_n}\sum_{k=1}^{n}(\xi_k-C_k). \tag{4}$$

容易想象，这问题决定于三个因素：

(i) $\{\xi_k\}$ 有什么性质？

(ii) 什么样的 $\{C_k\}$ 与 $\{D_k\}$?

(iii) 用哪种收敛性？

关于 (i), 我们以后总设 $\{\xi_k\}$ 是相互独立的 (非独立情形至今还研究得不完善).

当取 $C_k = E\xi_k$, $D_k = k$, 并且 ζ_n 几乎处处收敛 (或依概率收敛) 到 0 时，我们说，对 $\{\xi_k\}$ 强大数定理 (或大数定理) 成立，有时也说 $\{\xi_k\}$ 服从强大数定理 (或大数定理).

显然，如强大数定理成立，则大数定理必成立 (见 §3.1 定理 1).

当取 $C_k = E\xi_k$, $D_n = \sqrt{\sum\limits_{k=1}^{n} D\xi_k}$, 并且 ζ_n 的分布弱收敛到 $N(0,1)$ 正态分布时，我们就得到中心极限定理，留待下节详述.

(二) **大数定理**. 回忆在 §2.10 中所证明的车贝晓夫不等式：设随机变数 ξ 的方差有穷[1], 则对任意 $\varepsilon > 0$, 有

$$P(|\xi - E\xi| \geqslant \varepsilon) \leqslant \frac{D\xi}{\varepsilon^2}. \tag{5}$$

定理 1 设 $\{\xi_k\}$ 是随机变数列，对任一正整数 n, $D\Big(\sum\limits_{k=1}^{n} \xi_k\Big) < \infty$, 而且

$$\lim_{n\to\infty} \frac{1}{n^2} D\left(\sum_{k=1}^{n} \xi_k\right) = 0; \tag{6}$$

则 $\{\xi_k\}$ 服从大数定理，亦即对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k - E\xi_k)\right| \geqslant \varepsilon\right) = 0. \tag{7}$$

证 在 (5) 中取 $\xi = \frac{1}{n}\sum\limits_{k=1}^{n}(\xi_k - E\xi_k)$, 因而 $E\xi = 0$, 得

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k - E\xi_k)\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2 n^2} D\left(\sum_{k=1}^{n} \xi_k\right),$$

由此及 (6) 即得 (7). □

注意，如 $\{\xi_k\}$ 是独立随机变数列，则 (6) 化为

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} D\xi_k = 0. \tag{8}$$

系 1 (车贝晓夫) 设对独立随机变数列 $\{\xi_k\}$, 有常数 C, 使 $D\xi_k \leqslant C$, $k = 1, 2, \cdots$, 则 $\{\xi_k\}$ 服从大数定理.

证 此由

$$\frac{1}{n^2}\sum_{k=1}^{n} D\xi_k \leqslant \frac{nc}{n^2} \longrightarrow 0, \quad (n \to \infty) \tag{9}$$

[1] 根据 §2.9 中定义，如数学期望及方差存在，则必有穷，故这里 " 有穷 " 二字原可略去，但为了强调有穷性，我们仍把它明确写出.

及定理 1 推出. □

系 2 (贝努里) 设贝努里实验中，事件 A 每次出现的概率为 $p, 0 \leqslant p \leqslant 1$, 以 η_n 表前 n 次实验中 A 出现的次数，则对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{\eta_n}{n} - p\right| \geqslant \varepsilon\right) = 0. \tag{10}$$

证 考虑 (2) 中的 ξ_k, $D\xi_k = p(1-p) \leqslant 1$, 故 (10) 自系 1 推出. □

定理 2 (辛钦) 设 $\{\xi_k\}$ 是相互独立的随机变数列，有相同的分布，则 $\{\xi_k\}$ 服从大数定理的充要条件是 ξ_1 有有穷的数学期望.

证 以 $f(t)$ 表 ξ_k 的特征函数， $a = E\xi_k$, 根据同分布的假定，它们都不依赖于 k. 由 §2.11(一) 中 (iii),(iv),(v) 知 $\zeta_n = \frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 的特征函数是 $\left[f\left(\frac{t}{n}\right)\right]^n$; 而且当 $t \to 0$ 时有

$$f(t) = 1 + iat + o(t). \tag{11}$$

因此，对任一固定的 t, 有

$$\lim_{n\to\infty}\left[f\left(\frac{t}{n}\right)\right]^n = \lim_{n\to\infty}\left[1 + \frac{iat}{n} + o\left(\frac{t}{n}\right)\right]^n = e^{iat}. \tag{12}$$

但 e^{iat} 是单点分布函数

$$G(x) = \begin{cases} 1, & x \geqslant a, \\ 0, & x < a \end{cases} \tag{13}$$

的特征函数，故 (12) 说明 ζ_n 的分布函数 $G_n(x)$ 当 $n \to \infty$ 时弱收敛于 $G(x)$. 于是对任意 $\varepsilon > 0$ 有

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| \geqslant \varepsilon\right) &= P(\zeta_n \geqslant a + \varepsilon) + P(\zeta_n \leqslant a - \varepsilon) \\ &= 1 - G_n(a + \varepsilon - 0) + G_n(a - \varepsilon) \to 0 \quad (n \to \infty). \end{aligned} \tag{14}$$

因而条件的充分性得以证明. 必要性则是大数定理的定义中所要求的. □

显然，系 2 也可由定理 2 直接推出.

下面的定理 4 说明，把定理 2 中的“大数定理”四字改为“强大数定理”后，结果仍然正确.

(三) **柯尔莫哥洛夫不等式.** 强大数定理的证明要困难得多. 在这一段里先作些准备.

引理 1 设 $\xi_1, \cdots, \xi_n, \xi_{n+1}, \cdots, \xi_{n+m}$ 是独立随机变数，又 f_1 与 f_2 分别是 n 元及 m 元波雷耳可测函数，则随机变数 $f_1(\xi_1, \cdots, \xi_n)$ 与 $f_2(\xi_{n+1}, \cdots, \xi_{n+m})$ 相互独立.

证明完全可仿照 §2.7 系 1, 故从略.

在证明大数定理 (定理 1) 时，我们用到车贝晓夫不等式；为了证明强大数定理，需要一个更深刻的不等式 — 柯尔莫哥洛夫不等式 (即 (15) 式), 它可以看成车贝晓夫不等式的深化与推广，当 $n = 1$ 时，前者化为后者.

引理 2 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立随机变数，方差有穷，令 $s_n^2 = \sum\limits_{k=1}^{n} D\xi_k$, 则对任意 $\varepsilon > 0$, 有

$$P\left(\max_{1\leqslant k\leqslant n}\left|\sum_{j=1}^{k}(\xi_j - E\xi_j)\right| \geqslant \varepsilon\right) \leqslant \frac{s_n^2}{\varepsilon^2}. \tag{15}$$

在证明之前，先作些解释. 如 $s_n^2=0$, 则 $D\xi_j=0$, $P(\xi_j=E\xi_j)=1$ 因而 (15) 双方都化为 0 而 (15) 显然正确；故不妨设 $s_n^2>0$. 改写 (15) 为

$$P\left(\max_{1\leqslant k\leqslant n}\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geqslant \varepsilon s_n\right)\leqslant\frac{1}{\varepsilon^2}. \tag{16}$$

并简记 $A=\left(\max\limits_{1\leqslant k\leqslant n}\left|\sum\limits_{j=1}^{k}(\xi_j-E\xi_j)\right|\geqslant\varepsilon s_n\right)$, 我们有

$$A=\bigcup_{k=1}^{n}\left(\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geqslant\varepsilon s_n\right). \tag{17}$$

故 (16) 式表示： n 个事件 $\left(\left|\sum\limits_{j=1}^{k}(\xi_j-E\xi_j)\right|\geqslant\varepsilon s_n\right)$, $k=1,2,\cdots,n$ 中至少有一个出现的概率不大于 $1/\varepsilon^2$.(17) 式显然已把事件 A 分解为 n 个事件的和，但此 n 个事件未必不相交，为了要得到一个不相交的分解，对整数 v, $1\leqslant v\leqslant n$, 令

$$A_v=\left(\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|<\varepsilon s_n, k=0,\cdots,v-1;\left|\sum_{j=1}^{v}(\xi_j-E\xi_j)\right|\geqslant\varepsilon s_n\right); \tag{18}$$

换句话说， $\omega\in A_v$ 的充要条件是： n 个不等式

$$\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|<\varepsilon s_n,\quad (k=1,2,\cdots,n) \tag{19}$$

中，第一个不成立的是第 v 个，显然

$$A=\bigcup_{v=1}^{n}A_v;\quad A_uA_v=\phi\ \ (u\neq v). \tag{20}$$

故 (16) 式化为

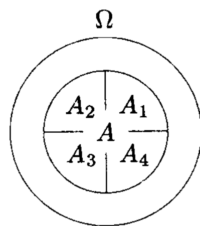
$$\sum_{v=1}^{n}P(A_v)\leqslant\frac{1}{\varepsilon^2}. \tag{21}$$

引理 2 之证　只要在 $s_n^2>0$ 的假设下证明 (21). 定义 n 个随机变数

$$y_v(\omega)=\begin{cases}1, & \text{如}\omega\in A_v,\\ 0, & \text{如}\omega\overline{\in}A_v.\end{cases}\quad (v=1,\cdots,n) \tag{22}$$

由于 $A_v(v=1,\cdots,n)$ 互不相交，每个 ω 至多只能属于一个 A_v, 故 $\sum\limits_{v=1}^{n}y_v(\omega)=0$ 或 1. 因而

$$\sum_{v=1}^{n}y_v(\omega)\leqslant 1.$$



(示意图， $n=4$)

令 $W_k=\sum\limits_{j=1}^{k}(\xi_j-E\xi_j)$, 以 W_n^2 乘上式两边，并取数学期望，得

$$\sum_{v=1}^{n}E(y_v\cdot W_n^2)\leqslant s_n^2. \tag{23}$$

为了估计 $E(y_v\cdot W_n^2)$, 注意

$$E(y_vW_n^2)=E(y_vW_v^2)+2E[y_v(W_n-W_v)W_v]+E[y_v(W_n-W_v)^2]. \tag{24}$$

但 $W_n-W_v=\sum\limits_{j=v+1}^{n}(\xi_j-E\xi_j)$ 只涉及 $\xi_{v+1},\cdots,\xi_n$; $y_vW_v=y_v\sum\limits_{j=1}^{v}(\xi_j-E\xi_j)$ 只涉及 $\xi_1,\cdots,\xi_v$, 既然 $\xi_1,\xi_2,\cdots,\xi_n$ 独立，故由引理 1, 知 W_n-W_v 与 y_vW_v 独立. 考虑到 $E(W_n-W_v)=0$, 得

$$E[y_v(W_n-W_v)W_v]=E(y_vW_v)E(W_n-W_v)=0. \tag{25}$$

又由 (22) 及 (18), 得 [1]

$$y_vW_v^2\geqslant y_v\varepsilon^2s_n^2. \tag{26}$$

综合 (23) (24) (25) (26), 并注意 $y_v\geqslant 0$, 即得

$$\begin{aligned}s_n^2&\geqslant\sum_{v=1}^{n}E(y_vW_n^2)\geqslant\sum_{v=1}^{n}E(y_vW_v^2)\geqslant\sum_{v=1}^{n}E(y_v\varepsilon^2s_n^2)\\&=\varepsilon^2s_n^2\sum_{v=1}^{n}Ey_v=\varepsilon^2s_n^2\sum_{v=1}^{n}P(A_v).\end{aligned}$$

由假设 $s_n^2>0$, 故 (21) 式得证. □

下面还要用到数学分析中三个简单事实，其中 2° 是周知的，为完全计，我们把证明都写出来.

1° 设 $\{c_k\}$ 为常数列，令 $s_n=\sum\limits_{k=1}^{n}c_k$, $n=1,2,\cdots$,

$$b_m=\sup\{|s_{m+k}-s_m|,\ k=1,2,\cdots\}. \tag{27}$$

$$b=\inf\{b_m,\ m=1,2,\cdots\}. \tag{28}$$

则级数 $\sum\limits_{k=1}^{\infty}c_k$ 收敛的充要条件是 $b=0$.

事实上，设 $\sum\limits_{k=1}^{\infty}c_k$ 收敛，则对任意 $\varepsilon>0$, 存在正整数 n_0, 使 $|s_{n_0+k}-s_{n_0}|<\varepsilon$, 对一切 $k=1,2,\cdots$ 成立，故 $0\leqslant b\leqslant b_{n_0}\leqslant\varepsilon$, 从而 $b=0$, 反之，由 $b=0$ 知对任意 $\varepsilon>0$, 存在正整数 n_1, 使 $b_{n_1}<\varepsilon$, 因而 $|s_{n_1+k}-s_{n_1}|<\varepsilon$ 对一切 $k=1,2,\cdots$ 成立，从而推知 $\sum\limits_{k=1}^{\infty}c_k$ 收敛.

2° 设 $\{c_k\}$ 为常数列，如 $\lim\limits_{k\to\infty}c_k=c$, 则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}c_k=c.$$

事实上，对 $n>n_0$, 有

$$\left|\frac{1}{n}\sum_{k=1}^{n}c_k-c\right|=\left|\frac{1}{n}\sum_{k=1}^{n}(c_k-c)\right|\leqslant\left|\frac{1}{n}\sum_{k=1}^{n_0}(c_k-c)\right|+\left|\frac{1}{n}\sum_{k=n_0+1}^{n}(c_k-c)\right|.$$

[1] 实际上，对任意 ω, 如 $\omega\overline{\in}A_v$, 由 (22) 知 (26) 双方在此 ω 上均化为 0, 故此时 (26) 成立；如 $\omega\in A_v, y_v(\omega)=1$, 再由 (18) 中最后一不等式即知此时 (26) 也成立.

对 $\varepsilon>0$, 选 n_0 使 $k>n_0$ 时，有 $|c_k-c|<\frac{\varepsilon}{2}$, 于是上式右方第二项小于 $\frac{n-n_0}{n}\frac{\varepsilon}{2}<\frac{\varepsilon}{2}$, 不论 $n(>n_0)$ 如何；右方第一项当 n 充分大后也小于 $\frac{\varepsilon}{2}$.

3° 如对实数列 $\{c_k\}$ 有 $\sum\limits_{k=1}^{\infty}\frac{c_k}{k}$ 收敛，则 $\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=1}^{n}c_k=0$.

事实上，令

$$s_0=0,\quad s_n=\sum_{k=1}^{n}\frac{c_k}{k},\quad t_n=\sum_{k=1}^{n}c_k,\quad n=1,2,\cdots.$$

则 $c_k=k(s_k-s_{k-1})$, $k=1,2,\cdots$, 故

$$t_{n+1}=\sum_{k=1}^{n+1}ks_k-\sum_{k=1}^{n+1}ks_{k-1}=-\sum_{k=1}^{n}s_k+(n+1)s_{n+1},\ n=1,2,\cdots,$$

$$\frac{t_{n+1}}{n+1}=-\frac{n}{n+1}\cdot\frac{1}{n}\sum_{k=1}^{n}s_k+s_{n+1}.$$

因 $\{s_n\}$ 收敛于有穷极限，由 2°, $\left\{\frac{1}{n}\sum\limits_{k=1}^{n}s_k\right\}$ 也收敛于同一极限，故

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}c_k=\lim_{n\to\infty}\frac{t_{n+1}}{n+1}=0.$$

(四) 强大数定理.

定理 3 (柯尔莫哥洛夫)　设 $\{\xi_k\}$ $k=1,2,\cdots$ 是独立随机变数列，且 $\sum\limits_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$, 则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\xi_k-E\xi_k)=0,\quad(\text{a.s.}).\tag{29}$$

证　定义

$$\xi_k'=\xi_k-E\xi_k$$

$$E\xi_k'=0,\ D\xi_k'=D\xi_k \text{ 仿 (27)(28), 令} s_n(\omega)=\sum_{k=1}^{n}\frac{\xi_k'(\omega)}{k},$$

$$b_m(\omega)=\sup\{|s_{m+k}(\omega)-s_m(\omega)|,\ k=1,2,\cdots\},$$

$$b(\omega)=\inf\{b_m(\omega),\ m=1,2,\cdots\}.$$

对任意实数 x,

$$(b_m(\omega)\leqslant x)=\bigcap_{k=1}^{\infty}(|s_{m+k}(\omega)-s_m(\omega)|\leqslant x)\in\mathcal{F}.$$

故 $b_m(\omega)$ 是随机变数，类似可知 $b(\omega)$ 也是随机变数.

对任意 $\varepsilon>0$ 及任二正整数 n,m, 由 (15) 得

$$P\Big(\max_{1\leqslant k\leqslant n}|s_{m+k}(\omega)-s_m(\omega)|\geqslant\varepsilon\Big)\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\xi_k'}{k^2}=\frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\xi_k}{k^2}.$$

令 $\eta_n(\omega)=\max\limits_{1\leqslant k\leqslant n}|s_{m+k}(\omega)-s_m(\omega)|$, 则 $\eta_n(\omega)\uparrow b_m(\omega)$, $(b_m(\omega)>\varepsilon)\subset\bigcup\limits_n(\eta_n(\omega)\geqslant\varepsilon)$, 由 §1.3 系 3 及上式得 $P(b_m(\omega)>\varepsilon)\leqslant\frac{1}{\varepsilon^2}\sum\limits_{k=m+1}^{\infty}\frac{D\xi_k}{k^2}$. 由此及 $b_m(\omega)\geqslant b(\omega)$ 推知

$$P(b(\omega)>\varepsilon)\leqslant P(b_m(\omega)>\varepsilon)\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{\infty}\frac{D\xi_k}{k^2}.$$

令 $m\to\infty$, 由于 $\sum\limits_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$, 得 $P(b(\omega)>\varepsilon)=0$. 从而自 $(b(\omega)>0)=\bigcup\limits_{n=1}^{\infty}\left(b(\omega)>\frac{1}{n}\right)$ 及 §1.3 系 3, 我们有

$$P(b(\omega)>0)=\lim_{n\to\infty}P\Big(b(\omega)>\frac{1}{n}\Big)=0,$$

于是几乎处处 $b(\omega)=0$. 由 1° 知 $\sum\limits_{k=1}^{\infty}\frac{D\xi'_k(\omega)}{k}$ 几乎处处收敛，再由 3° 得

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi'_k=0,\quad \text{(a.s.)}.$$

这就是 (29) 式. □

系 3 设对独立随机变数列 $\{\xi_k\}$, 有常数 c, 使 $D\xi_k\leqslant c$, $k=1,2,\cdots$, 则 $\{\xi_k\}$ 服从强大数定理.

证 由 $\sum\limits_{k=1}^{\infty}\frac{D\xi_k}{k^2}\leqslant c\sum\limits_{k=1}^{\infty}\frac{1}{k^2}<\infty$ 及定理 3 推出. □

系 4 (波雷耳) 设贝努里实验中，事件 A 每次出现的概率为 p, $0<p<1$, 以 η_n 表前 n 次实验中 A 出现的次数，则

$$\lim_{n\to\infty}\frac{\eta_n}{n}=p,\ \text{(a.s.)}\tag{30}$$

证 仿系 2 的证明，可见 (30) 自系 3 推出. □

定理 3 有着广泛的应用，但其中条件不是必要的. 在独立同分布情况，可以找到充要条件，为此先述一方法.

引理 3 (截尾法) 设 $\{\xi_k\}$, $k=1,2,\cdots$ 为具有有穷数学期望的随机变数列，有相同的分布函数 $F(x)$, 令

$$\xi_k^*=\begin{cases}\xi_k(\omega), & \text{如}\,|\xi_k(\omega)|\leqslant k,\\ 0, & \text{如}\,|\xi_k(\omega)|>k.\end{cases}\tag{31}$$

如果

$$\sum_{k=1}^{\infty}P(\xi_k^*(\omega)\neq\xi_k(\omega))<\infty,\tag{32}$$

而且 $\{\xi_k^*(\omega)\}$ 服从强大数定理，则 $\{\xi_k(\omega)\}$ 也服从强大数定理.

证 由 (32) 及 §3.1 引理 1 得 $P\Big(\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}[\xi_n^*(\omega)\neq\xi_n(\omega)]\Big)=0$, 换句话说，对几乎一切 ω,

$$\xi_k^*(\omega)\neq\xi_k(\omega),\ \text{只对有穷多个}\ k\ \text{成立}.\tag{33}$$

考虑

$$\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k(\omega)-E\xi_k)\right|\leqslant\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k(\omega)-\xi_k^*(\omega))\right|$$
$$+\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k^*(\omega)-E\xi_k^*)\right|+\left|\frac{1}{n}\sum_{k=1}^{n}(E\xi_k^*-E\xi_k)\right|. \tag{34}$$

根据 (33), 对几乎一切 ω, 右方第一项当 $n\to\infty$ 时趋于 0; 由假设 $\{\xi_n^*(\omega)\}$ 服从强大数定理, 故第二项也趋于 $0(n\to\infty)$; 今考虑第三项，由假定 $E\xi_1$ 有穷，故 $\int_{-\infty}^{\infty}|x|dF(x)<\infty$, 从而若令

$$c_k=\left|\int_{|x|>k}xdF(x)\right|,$$

则

$$\lim_{k\to\infty}c_k\leqslant\lim_{k\to\infty}\int_{|x|>k}|x|\,dF(x)=0. \tag{35}$$

令

$$g_k(x)=\begin{cases}x, & \text{如}|x|\leqslant k,\\ 0, & \text{如}|x|>k,\end{cases}$$

则 $g_k(x)$ 是波雷耳可测函数，而且 $\xi_k^*=g_k(\xi_k)$. 由于

$$E\xi_k^*=\int_{-\infty}^{\infty}g_k(x)dF(x)=\int_{|x|\leqslant k}xdF(x),$$

故对 $n>N$ 有

$$\left|\frac{1}{n}\sum_{k=1}^{n}(E\xi_k^*-E\xi_k)\right|\leqslant\frac{1}{n}\sum_{k=1}^{n}|E\xi_k^*-E\xi_k|$$
$$=\frac{1}{n}\sum_{k=1}^{n}\left|\int_{|x|>k}xdF(x)\right|=\frac{1}{n}\sum_{k=1}^{n}c_k.$$

由 (35) 及上面的事实 2°, 即知 (34) 中右方第三项也趋于 0, 从而 $\{\xi_k(\omega)\}$ 服从强大数定理.
□

引理 3 将对 $\{\xi_k\}$ 的研究化为对 $\{\xi_k^*\}$ 的研究，由于每个 ξ_k^* 有界，故带来不少方便.

定理 4 设 $\{\xi_k\}$ $k=1,2,\cdots$ 是独立随机变数列，有相同的分布函数 $F(x)$, 则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\xi_k-E\xi_k)=0,\quad \text{(a.s.)} \tag{36}$$

成立的充要条件是 ξ_1 有有穷的数学期望.

记 $a=E\xi_1$, 由于 $\{\xi_k\}$ 同分布， $E\xi_k=a$, 故 (36) 可写为

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k=a,\quad \text{(a.s.)}.$$

证 必要性是强大数定理成立的定义所要求的；下证充分性，对定理中的 $\{\xi_k\}$ 用 (31) 定义 $\{\xi_k^*\}$，根据 §2.7 系 1 知 $\{\xi_k^*\}$ 也是独立随机变数列. 因 $E\xi_1$ 有穷，故 $\int_{-\infty}^{\infty}|x|dF(x)<\infty$. 由同分布的假定

$$
\begin{aligned}
\sum_{k=1}^{\infty}P(\xi_k^*\neq\xi_k)&=\sum_{k=1}^{\infty}P(|\xi_k|>k)=\sum_{k=1}^{\infty}P(|\xi_1|>k)\\
&=\sum_{k=1}^{\infty}\sum_{l=k}^{\infty}P(l<|\xi_1|\leqslant l+1)=\sum_{l=1}^{\infty}lP(l<|\xi_1|\leqslant l+1)\\
&\leqslant\sum_{l=0}^{\infty}\int_{l<|x|\leqslant l+1}|x|dF(x)\leqslant\int_{-\infty}^{\infty}|x|dF(x)<\infty. \qquad (37)
\end{aligned}
$$

因而根据引理 3，只要证明 $\{\xi_k^*\}$ 服从强大数定理. 为此，由定理 3，只需验证 $\sum\limits_{k=1}^{\infty}\frac{D\xi_k^*}{k^2}<\infty$. 注意 [1]

$$
\begin{aligned}
D\xi_k^*\leqslant E(\xi_k^*)^2&=\int_{|x|\leqslant k}x^2dF(x)\\
&\leqslant\sum_{l=0}^{k}(l+1)^2P(l\leqslant|\xi_1|<l+1);\\
\sum_{k=1}^{\infty}\frac{D\xi_k^*}{k^2}&\leqslant\sum_{k=1}^{\infty}\sum_{l=0}^{k}\frac{(l+1)^2}{k^2}P(l\leqslant|\xi|<l+1)\\
&\leqslant\sum_{l=0}^{\infty}P(l\leqslant|\xi|<l+1)(l+1)^2\sum_{k=l}^{\infty}\frac{1}{k^2} \qquad (38)
\end{aligned}
$$

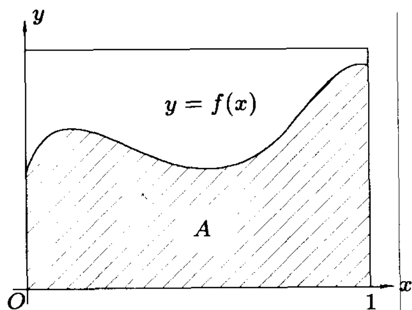
利用 $\sum\limits_{k=l}^{\infty}\frac{1}{(k+1)^2}<\sum\limits_{k=l}^{\infty}\int_k^{k+1}\frac{dx}{x^2}=\int_l^{\infty}\frac{dx}{x^2}=\frac{1}{l}$，得

$$
\sum_{k=l}^{\infty}\frac{1}{k^2}<\frac{1}{l^2}+\frac{1}{l}<\frac{2}{l},
$$

故由 (38) 及 (37) 中的推导得 $\sum\limits_{k=l}^{\infty}\frac{D\xi_k^*}{k^2}<\infty$. □

(五) 例. 下二例说明大数定理的一些应用.

例 1 设 $f(x)$ $(a\leqslant x\leqslant b)$ 是连续函数，试用概率方法来近似计算积分 $\int_a^b f(x)\,dx$.



设 $|f(x)|$ 的一上界为 $k(>0)$，$f(x)$ 的下确界为 h，则 $0\leqslant\frac{f(x)-h}{2k}\leqslant 1$，故不妨假定 $0\leqslant f(x)\leqslant 1$. 引进新变量 z：$x=(b-a)z+a$ 后，可将 x 轴上的区间 $[a,b]$ 变为 z 轴上的 $[0,1]$，故不妨设 $a=0$，$b=1$.

考虑几何型随机试验 E：向矩形 $0\leqslant x\leqslant 1$，$0\leqslant y\leqslant 1$ 中均匀分布地掷点. 将 E 独立地重复作下去，以 A 表此矩形中曲线 $y=f(x)$ 下的区域，即

[1] 但当 $l=0$ 时，$\sum_{k=l}^{\infty}\frac{1}{k^2}$ 应理解为 $\sum_{k=1}^{\infty}\frac{1}{k^2}$.

$$A = ((x, y):\ 0 \leqslant y \leqslant f(x),\ x \in [0,1]).$$

并定义随机变数 ξ_k:

$$\xi_k = \begin{cases} 1, & \text{如第 } k \text{ 次掷的点落于 } A \text{ 中}, \\ 0, & \text{反之}. \end{cases}$$

则 $\{\xi_k\}$ $k = 1, 2, \cdots$ 独立同分布，而且

$$E\xi_k = P(\xi_k = 1) = |A| = \int_0^1 f(x)\,dx,$$

$|A|$ 表 A 的面积. 由系 4

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \xi_k = \int_0^1 f(x)\,dx, \quad \text{(a.s.)}.$$

这表示当 n 充分大时，前 n 次试验中落于 A 中的点数 $\sum\limits_{k=1}^{n} \xi_k$ 除以 n 后以任意接近于 1 的概率与 $\int_0^1 f(x)\,dx$ 近似.

这种近似计算法也叫 Monte-Carlo 方法. 进一步的讨论见第七章.

例 2 在函数逼近论中有著名的

外尔斯特拉斯 (Weierstrass) **定理**： 设 $f(x)$ 为闭区间 $a \leqslant x \leqslant b$ 中的任一连续函数，则必存在多项式序列 $\{Q_n(x)\}$, 均匀收敛于 $f(x)$, $x \in [a, b]$.

试给出这定理的一简单概率证明. 通过一线性变换 (参看上例), 可把 $[a, b]$ 变为 $[0, 1]$, 故不妨设 $a = 0, b = 1$, 令

$$Q_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f\left(\frac{m}{n}\right). \tag{39}$$

显然， $Q_n(0) = f(0), Q_n(1) = f(1)$, 故只要考虑 $(0,1)$ 中的 x. 任取一贝努里实验 $\widetilde{E} = E^{\infty}$, 使事件 A 在每次试验 E 中出现的概率恰为 x, x 任意固定；并以 η_n 表前 n 次试验中 A 出现的总次数，则由 §2.9(24) 得

$$\begin{aligned} f(x) - Q_n(x) = &\sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} \left[f(x) - f\left(\frac{m}{n}\right)\right] = E\left[f(x) - f\left(\frac{\eta_n}{n}\right)\right] \\ = &P\left(\left|\frac{\eta_n}{n} - x\right| < \delta\right) E\left[f(x) - f\left(\frac{\eta_n}{n}\right)\right] \,\Big|\, \left|\frac{\eta_n}{n} - x\right| < \delta\Big) \\ &+ P\left(\left|\frac{\eta_n}{n} - x\right| \geqslant \delta\right) E\left[f(x) - f\left(\frac{\eta_n}{n}\right)\right] \,\Big|\, \left|\frac{\eta_n}{n} - x\right| \geqslant \delta\Big), \end{aligned} \tag{40}$$

其中 $\delta > 0$ 为如下选定的数：由 $f(x)$ 的连续性，对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使当 $|x - y| < \delta$, $0 \leqslant x, y \leqslant 1$ 时，有

$$|f(x) - f(y)| < \frac{\varepsilon}{2}. \tag{41}$$

令 $M = \sup\limits_{0 \leqslant x \leqslant 1} |f(x)|$, 由 (40)(41) 得

$$|f(x) - Q_n(x)| \leqslant 1 \cdot \frac{\varepsilon}{2} + P\left(\left|\frac{\eta_n}{n} - x\right| \geqslant \delta\right) \cdot 2M. \tag{42}$$

令 $n\to\infty$, 由大数定理， $\lim\limits_{n\to\infty}P\left(\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right)=0$, 从而得证

$$\lim_{n\to\infty}Q_n(x)=f(x),\ (0\leqslant x\leqslant 1).$$

为证上式中收敛的均匀性，利用车贝晓夫不等式.

$$P\left(\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right)\leqslant\frac{D\eta_n}{n^2\delta^2}=\frac{nx(1-x)}{n^2\delta^2}<\frac{1}{n\delta^2}.$$

故当 $n>\frac{4M}{\varepsilon\delta^2}$ 时，由上式及 (42) 立得 $\sup\limits_{0\leqslant x\leqslant 1}|f(x)-Q_n(x)|<\varepsilon$. □

§3.4 中心极限定理

(一) 问题的直观背景. 在实际中，常常需要考虑许多随机因素所产生的总影响.

例 1 设炮弹射击的目标位置是原点 (0,0), 炮弹的落点设为 (X,Y), 它的一个坐标，例如 X, 也是落点对目标沿 $x-$ 轴的偏差， X(或 Y) 是随机的，产生偏差的原因有种种：瞄准时有误差 ξ_1, 炮弹或炮身结构所引起的误差 ξ_2, 空气阻力产生的误差 $\xi_3,\cdots$. 因而， X 可看成为这些误差的总和： $X=\sum\limits_i\xi_i$. 对我们重要的是研究 X 的分布.

例 2 再看一个熟悉的例. 考虑可列重贝努里试验 $\widetilde{E}=E^\infty$, 事件在每次试验 E 中出现的概率为 $p, 0<p<1$, 不出现的概率为 $q=1-p$. 令 $\{\xi_k\}$ 为连系于 $\widetilde{E}$ 的随机变数列，即

$$\xi_k=\begin{cases}1, & \text{如 } A \text{ 在第 } k \text{ 次试验中出现},\\ 0, & \text{反之}.\end{cases}\tag{1}$$

因而 $\sum\limits_{k=1}^n\xi_k$ 是前 n 次试验中 A 出现的总次数，我们关心的问题是：当 $n\to\infty$ 时， $\sum\limits_{k=1}^n\xi_k$ 的分布函数趋于什么？由于 $\lim\limits_{n\to\infty}\sum\limits_{k=1}^n\xi_k$ 可以取 ∞ 为值，故最好不研究 $\sum\limits_{k=1}^n\xi_k$ 本身而考虑 (譬如说) 它的标准化随机变数

$$\zeta_n=\frac{\sum\limits_{k=1}^n(\xi_k-E\xi_k)}{\sqrt{D\left(\sum\limits_{k=1}^n\xi_k\right)}}\tag{2}$$

的分布函数的极限.

为了求出这个极限分布，我们先作些具体的，直观的分析. 设

$$P(\xi_k=1)=\frac{2}{3},\ \ P(\xi_k=0)=\frac{1}{3}.$$

下在画出 $\sum\limits_{k=1}^n\xi_k$ $(n=1,2,4,8,32)$ 的密度图 (见图 1). 由此可见，当 $n\to\infty$ 时， $\sum\limits_{k=1}^n\xi_k$ 的密度趋于正态分布密度. 为了防止这些密度的中心 (即数学期望) 趋向 ∞ 以及方差无限增大，我们不考虑 $\sum\limits_{k=1}^n\xi_k$ 而考虑 (2) 中的 ζ_n.

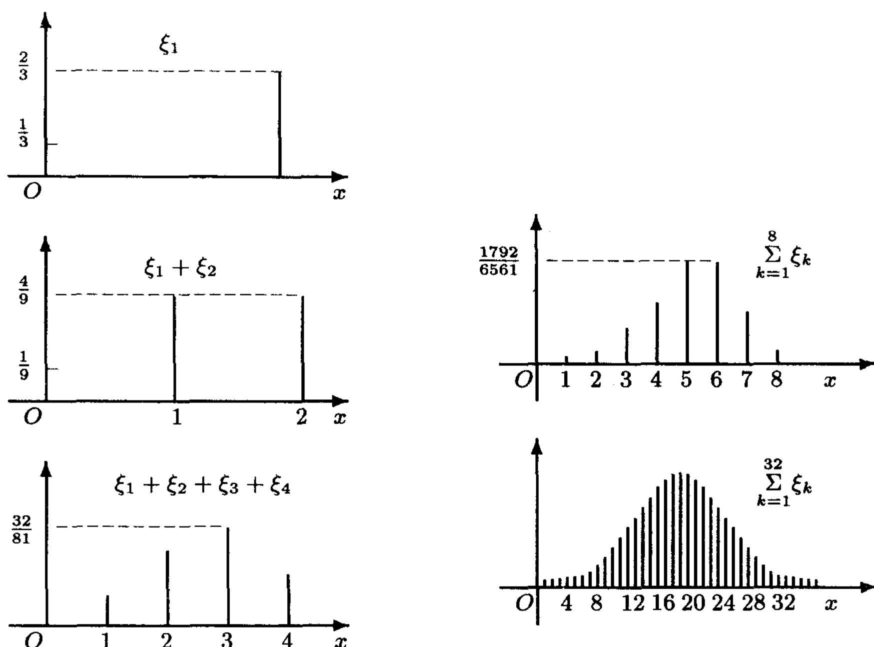


图 1

当 ξ_k 有 $(0,1)$ 均匀分布时，类似的图请参看 §6.1 图 2.

(二) **同分布情形**. 设 $\xi_k(k=1,2,\cdots)$ 的方差 $D\xi_k$ 有穷，大于 0, 令

$$a_k = E\xi_k,\quad b_k^2 = D\xi_k,\quad B_n^2 = \sum_{k=1}^{n} b_k^2. \tag{3}$$

我们说，随机变数列 $\{\xi_k\}$ 服从中心极限定理，如果关于 $x \in R_1$ 均匀地有

$$\lim_{n\to\infty} P\left(\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k - a_k) \leqslant x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz. \tag{4}$$

(4) 式表示：随机变数 $\frac{1}{B_n}\sum\limits_{k=1}^{n}(\xi_k - a_k)$ 的分布函数关于 x 均匀地趋于正态分布 $N(0,1)$ 的分布函数. 我们的主要目的是：研究在什么条件下，$\{\xi_k\}$ 服从中心极限定理，一个简单而又常用的结果是：

定理 1 设独立随机变数列 $\{\xi_k\}$ 有相同的分布，而且 $0 < D\xi_k < \infty$, 则 (4) 式关于 x 均匀成立.

证 由于分布相同，$a_k = a_1, b_k^2 = b_1^2$,(4) 式化为 [1]

---

[1] 在定理 1 的条件下，如再补设 $\beta_3 = E|\xi_r - E\xi_r|^3 < \infty$, 则

$$\max_x \left| P\Big(\frac{1}{b_1\sqrt{n}}\sum_{k=1}^{n}(\xi_k - a_1) \leqslant x\Big) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}dz \right| < \frac{c}{\sqrt{n}}\frac{\beta_3}{b_1^3},$$

其中 c 为某常数，已知 $c \geqslant \frac{1}{\sqrt{2\pi}}$, 但其精确值还未求出. Золотарев 于 1967 年得到一较好的结果是 $c < 0.82$(参看 [38],136 页；[49],159 页). 上式对误差估计有用. 对不同分布情况也有类似的结果，见 [49].

$$\lim_{n\to\infty} P\left(\frac{1}{b_1\sqrt{n}}\sum_{k=1}^{n}(\xi_k - a_1) \leqslant x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz. \tag{5}$$

以 $f_k(t)$ 表 $\xi_k - a_1$ 的特征函数，因 $\xi_k - a_1 (k = 1, 2, \cdots)$ 也同分布，故 $f_k(t) = f_1(t)$ 不依赖于 k. 根据 §2.11(8),(10), 可见 $\frac{1}{b_1\sqrt{n}}\sum\limits_{k=1}^{n}(\xi_k - a_1)$ 的特征函数 $\varphi_n(t)$ 为

$$\varphi_n(t) = \left[f\left(\frac{t}{b_1\sqrt{n}}\right)\right]^n. \tag{6}$$

由于 $\xi_k - a_1$ 的前两阶矩为 0 及 b_1^2, 由 §2.11(12), 在 $t = 0$ 附近可将 $f(t)$ 按泰勒公式展开为

$$f(t) = 1 - \frac{1}{2}b_1^2 t^2 + o(t^2). \tag{7}$$

以 $\frac{t}{b_1\sqrt{n}}$ 代 t, 并利用 (6), 得

$$\varphi_n(t) = \left[1 - \frac{t^2}{2n} + \frac{o(t^2)}{n}\right]^n \to e^{-\frac{t^2}{2}}\ (n \to \infty).$$

注意 $e^{-\frac{t^2}{2}}$ 是 $N(0,1)$ 分布的特征函数，由上式及 §3.2 定理 3, 即得证 (5) 式对每 $x \in (-\infty, \infty)$ 成立. 根据 §3.1 定理 3,(5) 中的收敛对 x 还是均匀的. □

系 1 (隶莫佛 - 拉普拉斯 [de Moivre-Laplace]) 以 η_n 表贝努里 n 次试验中事件 A 出现的总次数，每次试验中 A 出现的概率为 p, $0 < p < 1$, 则当 $n \to \infty$ 时，对 $a, b (a < b)$ 均匀地有

$$\lim_{n\to\infty} P\left(a < \frac{\eta_n - np}{\sqrt{npq}} \leqslant b\right) = \frac{1}{\sqrt{2\pi}}\int_{a}^{b} e^{-\frac{z^2}{2}}\,dz \tag{8}$$

其中 $q = 1 - p$.

证 用 (1) 定义 ξ_k, $\{\xi_k\}$ 是独立随机变数列，有相同的分布密度矩阵 $\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$ 则

$$\eta_n = \sum_{k=1}^{n}\xi_k,\quad E\xi_k = p,\quad D\xi_k = pq > 0,\quad B_n^2 = \sum_{k=1}^{n} D\xi_k = npq.$$

故由定理 1 立得 (8). □

例 3 对系 1 中的贝努里试验，试求 $P(\alpha < \eta_n \leqslant \beta)$ 的值 $(\alpha < \beta)$.

解 由于 η_n 有二项分布 $B(n, p)$, 故

$$P(\alpha < \eta_n \leqslant \beta) = \sum_{\alpha < k \leqslant \beta}\binom{n}{k}p^k q^{n-k}. \tag{9}$$

当 n 甚大时，如 §2.4 所指出，上值的计算很繁，利用 (8) 可求出 $P(\alpha < \eta_n \leqslant \beta)$ 的近似值：

$$\begin{aligned} P(\alpha < \eta_n \leqslant \beta) &= P\left(\frac{\alpha - np}{\sqrt{npq}} < \frac{\eta_n - np}{\sqrt{npq}} \leqslant \frac{\beta - np}{\sqrt{npq}}\right) \\ &\approx \frac{1}{\sqrt{2\pi}}\int_{\frac{\alpha - np}{\sqrt{npq}}}^{\frac{\beta - np}{\sqrt{npq}}} e^{-\frac{z^2}{2}}\,dz = \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right), \end{aligned} \tag{10}$$

其中

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}\,dz. \tag{11}$$

查表即可容易地算出 $\Phi\Big(\frac{\beta-np}{\sqrt{npq}}\Big)-\Phi\Big(\frac{\alpha-np}{\sqrt{npq}}\Big)$. 由 (9) 可见，此值也是 $\sum\limits_{\alpha<k\leqslant\beta}\binom{n}{k}p^kq^{n-k}$ 的近似值. 因此，可以用正态分布来近似计算二项分布的值.

例 4 射击不断地独立进行，设每次射中的概率为 $\frac{1}{10}$.

(i) 试求 500 次射击中，射中的次数在区间 $(49,55]$ 之中的概率 P_1.

(ii) 问最少要射击多少次，才能使射中的次数超过 50 次的概率大于已给正数 q?

解 以 η_n 表前 n 次射击总射中的次数 [1].

(i) $np=500\cdot\frac{1}{10}=50$, $npq=500\cdot\frac{1}{10}\cdot\frac{9}{10}=45$, $\alpha=49$, $\beta=55$.

$$\begin{aligned}P_1=P(49<\eta_{500}\leqslant 55)&\approx\Phi\Big(\frac{55-50}{\sqrt{45}}\Big)-\Phi\Big(\frac{49-50}{\sqrt{45}}\Big)\\&=\Phi\Big(\frac{5}{\sqrt{45}}\Big)-\Phi\Big(\frac{-1}{\sqrt{45}}\Big)=0.323\cdots.\end{aligned}$$

(ii) 所需最少射击次数是满足不等式

$$P(50<\eta_n)>q$$

的最小正整数 n, 在 (10) 中令 $\beta\to\infty$, 得

$$P(50<\eta_n)=P\left(\frac{50-n\cdot\frac{1}{10}}{\sqrt{n\cdot\frac{1}{10}\cdot\frac{9}{10}}}<\frac{\eta_n-n\cdot\frac{1}{10}}{\sqrt{n\cdot\frac{1}{10}\cdot\frac{9}{10}}}\right)\approx 1-\Phi\Big(\frac{500-n}{3\sqrt{n}}\Big). \tag{12}$$

故自 $1-\Phi\Big(\frac{500-n}{3\sqrt{n}}\Big)>q$ 解出最小的正整数 n 即为所求次数的近似值.

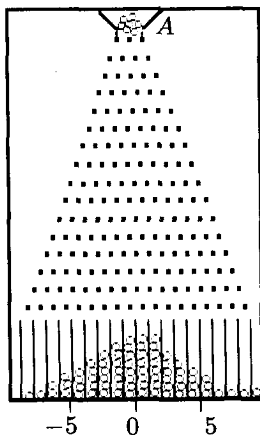


图 2

例 5 (钉板实验) 图 2 中每一黑点代表钉在板上的一颗钉子，它们间的距离相等，上面一颗恰巧在下面两颗的正中间. 从入口 A 处放进一个小圆球，它的直径略小于二钉间的距离，由于板是倾斜放着的，球每碰到一次钉子，就以概率 $\frac{1}{2}$ 滚向左下边 (或右下边), 于是又碰到下一个钉子，如此继续，直到滚入板底的一个格子内为止. 把许多小球从 A 放下去，它们在板底所堆成的曲线就近似于 $N(0,\sqrt{n})$ 正态分布的密度曲线，只要球的个数充分地大，这里 n 是钉子的横排排数，这个实验是由 Galton 设计的.

现在用定理 1 来解释这个现象. 定义

$$\xi_k=\begin{cases}1, & \text{如第 } k \text{ 次碰钉后小球向右;}\\ -1, & \text{如第次碰钉后小球向左.}\end{cases}$$

[1] 如要近似计算 $P(\eta_n=k)$, k 为非负整数，则可利用 $P(\eta_n=k)=P\big(k-\frac{1}{2}<\eta_n\leqslant k+\frac{1}{2}\big)$, 然后仿下面的 (i) 计算 (更详细的讨论见 [15],172 页).

于是 ξ_k 的密度矩阵为 $\begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, 而 $\eta_n = \sum\limits_{k=1}^{n} \xi_k$ 标示第 n 次碰钉后球的位置 [1]. 由于

$$a_k = E\xi_k = 0, \qquad b_k = D\xi_k = 1,$$

根据 (5), 得

$$\lim_{n\to\infty} P\Big(\frac{\eta_n}{\sqrt{n}} \leqslant x\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz.$$

这表示 $\frac{\eta_n}{\sqrt{n}}$ 的分布当 n 充分大时近似于 $N(0,1)$, 亦即 η_n 的分布近似于 $N(0,\sqrt{n})$. 由上式得

$$P(-l \leqslant \eta_n \leqslant l) = P\Big(\frac{-l}{\sqrt{n}} \leqslant \frac{\eta_n}{\sqrt{n}} \leqslant \frac{l}{\sqrt{n}}\Big) \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{-l}{\sqrt{n}}}^{\frac{l}{\sqrt{n}}} e^{-\frac{z^2}{2}}\,dz.$$

今设 $n=16$, 则

$$P(-l \leqslant \eta_{16} \leqslant l) \approx \frac{1}{\sqrt{2\pi}} \int_{-\frac{l}{4}}^{\frac{l}{4}} e^{-\frac{z^2}{2}}\,dz$$

右方值可由正态表查出，例如

$$P(-1 \leqslant \eta_{16} \leqslant 1) \approx \frac{1}{\sqrt{2\pi}} \int_{-0.25}^{0.25} e^{-\frac{z^2}{2}}\,dz = 0.1974.$$

现在独立地投掷 60 个小球，则大约有 $60\times 0.1974 \approx 12$ 个落在 $[-1,1]$ 两格之中. 列表如下：

区　　间	近似概率	近似球数	区　　间	近似概率	近似球数
$[-1,1]$	0.1974	12	$[-6,6]$	0.8664	52
$[-2,2]$	0.3829	23	$[-7,7]$	0.9199	55
$[-3,3]$	0.5467	33	$[-8,8]$	0.9545	57
$[-4,4]$	0.6827	41	$[-9,9]$	0.9756	59
$[-5,5]$	0.7887	47	$[-10,10]$	0.9876	60

表中数据说明：一个球在碰过 16 次钉子后落于 $[-2,2]$ 中的概率为 0.3829; 如果将 60 个球做这实验，则约有 23 个球落在 $[-2,2]$ 之中.

(三) **一般情形**. 定理 1 中 "$\{\xi_k\}$ 有相同分布" 的假定太强，现在来放宽这一假定而引进著名的

麟德伯尔格 (Lindeberg) 条件：对任意 $\tau > 0$, 有

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 dF_k(x) = 0; \tag{13}$$

其中 $F_k(x)$ 是 ξ_k 的分布函数， a_k, B_n 由 (3) 定义.

先看这一条件有什么概率意义. 引进事件

$$A_k = (|\xi_k - a_k| > \tau B_n), \quad k = 1,2,\cdots,n, \tag{14}$$

[1] 容易看出，这问题与下列随机徘徊问题等价：设质点 A 在整数点上运动，如果它于时刻 k 位于点 j, 则下一时刻转移到点 $j+1$ 或 $j-1$, 概率分别为 $\frac{1}{2}$, 而且与以前的转移独立. 试研究自 0 点出发，经过 n 次转移后，质点所在位置 η_n 的分布. 这是一种所谓的马尔科夫链的特例.

我们有

$$P(\max_{1\leqslant k\leqslant n}|\xi_k-a_k|>\tau B_n)=P\Big(\bigcup_{k=1}^{n}A_k\Big)\leqslant\sum_{k=1}^{n}P(A_k)$$
$$=\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}dF_k(x)\leqslant\frac{1}{(\tau B_n)^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x).$$

由此可见，麟德伯尔格条件保证：对任意 $\tau>0$,

$$\lim_{n\to\infty}P\Big(\max_{1\leqslant k\leqslant n}\Big|\frac{\xi_k-a_k}{B_n}\Big|>\tau\Big)=0. \tag{15}$$

粗略地说，上式表示：当 n 充分大后，参与构成 (5) 中左方总和 $\frac{1}{B_n}\sum\limits_{k=1}^{n}(\xi_k-a_k)$ 的每一被加项 $\frac{\xi_k-a_k}{B_n}$ 要依概率 "均匀地小"，因而任一被加项对总和的极限分布不会产生显著的影响.

容易验证，在定理 1 的假定下，麟德伯尔格条件满足. 实际上，由于这时 $F_k(x)=F(x)$, $a_k=a$, $b_k^2=b^2$ 都与 k 无关，又 $B_n^2=nb^2$, 故

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)=\frac{1}{nb^2}\cdot n\int_{|x-a|>\tau b\sqrt{n}}(x-a)^2dF(x); \tag{16}$$

既然方差 $D\xi_1=\int_{-\infty}^{\infty}(x-a)^2dF(x)<\infty$, 我们有

$$\lim_{n\to\infty}\int_{|x-a|>\tau b\sqrt{n}}(x-a)^2dF(x)=0;$$

由此及 (16) 知 (13) 成立.

定理 2 (麟德伯尔格) 设独立随机变数列 $\{\xi_k\}$ $k=0,1,2,\cdots$ 满足麟德伯尔格条件，则 (4) 式关于 x 均匀成立.

为证此，先证下列三个不等式：对任意实数 a, 有

$$|e^{ia}-1|\leqslant|a|; \tag{17}$$
$$|e^{ia}-1-ia|\leqslant\frac{a^2}{2!}; \tag{18}$$
$$\Big|e^{ia}-1-ia+\frac{a^2}{2}\Big|\leqslant\frac{|a|^3}{3!}; \tag{19}$$

实际上，对 $a=0$ 上三式明显. 设 $a>0$, 则

$$|e^{ia}-1|=\Big|\int_0^a e^{ix}\,dx\Big|\leqslant a;$$
$$|e^{ia}-1-ia|=\Big|\int_0^a(e^{ix}-1)\,dx\Big|\leqslant\int_0^a xdx=\frac{a^2}{2!};$$
$$\Big|e^{ia}-1-ia+\frac{a^2}{2}\Big|=\Big|\int_0^a(e^{ix}-1-ix)\,dx\Big|$$
$$\leqslant\int_0^a|e^{ix}-1-ix|\,dx\leqslant\int_0^a\frac{x^2}{2!}\,dx=\frac{a^3}{3!}. \tag{20}$$

利用 $e^{ia}=\cos a+i\sin a$, 可见 (17),(18),(19) 两方都是 a 的偶函数，故它们对 $a<0$ 也成立.

定理 2 的证 先把记号简化. 令

$$\xi_{nk}=\frac{\xi_k-a_k}{B_n}. \tag{21}$$

以 $f_{nk}(t)$, $F_{nk}(x)$ 分别表 ξ_{nk} 的特征函数与分布函数，因而

$$F_{nk}(x)=P(\xi_k\leqslant B_nx+a_k)=F_k(B_nx+a_k), \tag{22}$$

$$E\xi_{nk}=\int_{-\infty}^{\infty}xdF_{nk}(x)=0,\quad D\xi_{nk}=\frac{D\xi_k}{B_n^2}, \tag{23}$$

$$\sum_{k=1}^{n}D\xi_{nk}=\sum_{k=1}^{n}\int_{-\infty}^{\infty}x^2dF_{nk}(x)=\frac{1}{B_n^2}\sum_{k=1}^{n}D\xi_k=1. \tag{24}$$

在这些记号下，由 (22)

$$\frac{1}{B_n^2}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)=\int_{|\frac{x-a_k}{B_n}|>\tau}\left(\frac{x-a_k}{B_n}\right)^2dF_k(x)=\int_{|y|>\tau}y^2dF_{nk}(y).$$

故 (13) 式化为：对任意 $\tau>0$,

$$\lim_{n\to\infty}\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x)=0; \tag{25}$$

而 (4) 式化为：对 τ 均匀地有

$$\lim_{n\to\infty}P\left(\sum_{k=1}^{n}\xi_{nk}\leqslant x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}dz. \tag{26}$$

如果在条件 (25) 下，能够证明 $\sum\limits_{k=1}^{n}\xi_{nk}$ 的特征函数

$$\varphi_n(t)=\prod_{k=1}^{n}f_{nk}(t)\to e^{-\frac{t^2}{2}}\ (n\to\infty).$$

(回忆 $e^{-\frac{t^2}{2}}$ 为 $N(0,1)$ 的特征函数), 亦即

$$\log\varphi_n(t)=\sum_{k=1}^{n}\log f_{nk}(t)\to-\frac{t^2}{2},\ (n\to\infty) \tag{27}$$

(对数取主值), 那么根据 §3.2 定理 3,(26) 式成立；再由 §3.1 定理 3,(26) 中收敛对 $x\in R_1$ 还是均匀的，于是定理 2 得以证明.

为了证明 (27), 分成两步.

(甲) 先证 $\log\varphi_n(t)$ 可展开为

$$\log\varphi_n(t)=\sum_{k=1}^{n}(f_{nk}(t)-1)+R_n(t), \tag{28}$$

其中函数 $R_n(t)$ 在任意有穷 t 区间内均匀趋于 0.

实际上，由 (23) 中前一式

$$f_{nk}(t)-1=\int_{-\infty}^{\infty}(e^{itx}-1-itx)dF_{nk}(x). \tag{29}$$

根据 (18)

$$\begin{aligned}|f_{nk}(t)-1| &\leqslant \frac{t^2}{2}\int_{-\infty}^{\infty}x^2dF_{nk}(x)=\frac{t^2}{2}\left[\int_{|x|\leqslant\varepsilon}x^2dF_{nk}(x)+\int_{|x|>\varepsilon}x^2dF_{nk}(x)\right]\\ &\leqslant \frac{t^2}{2}\left[\varepsilon^2+\int_{|x|>\varepsilon}x^2dF_{nk}(x)\right],\end{aligned} \tag{30}$$

其中 $\varepsilon>0$ 任意. 由 (25), 对一切充分大的 n 有 $\int_{|x|>\varepsilon}x^2dF_{nk}(x)<\varepsilon^2(1\leqslant k\leqslant n)$; 从而关于 $k(1\leqslant k\leqslant n)$ 及任何有限区间 $[-T,T]$ 中的 t, 同时有

$$|f_{nk}(t)-1|\leqslant\varepsilon^2T^2;\quad \max_{1\leqslant k\leqslant n}|f_{nk}(t)-1|\leqslant\varepsilon^2T^2.$$

因而对任意 $t\in[-T,T]$, 均匀地有

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant n}|f_{nk}(t)-1|=0. \tag{31}$$

特别，当 $t\in[-T,T]$ 时，对一切充分大的 n, 下式成立：

$$|f_{nk}(t)-1|<\frac{1}{2}. \tag{32}$$

因此，在 $[-T,T]$ 中，有展开式

$$\begin{aligned}\log\varphi_n(t)&=\sum_{k=1}^{n}\log f_{nk}(t)=\sum_{k=1}^{n}\log[1+(f_{nk}(t)-1)]\\ &=\sum_{k=1}^{n}(f_{nk}(t)-1)+R_n(t).\end{aligned} \tag{33}$$

其中

$$R_n(t)=\sum_{k=1}^{n}\sum_{s=2}^{\infty}\frac{(-1)^{s-1}}{s}(f_{nk}(t)-1)^s.$$

由 (32)

$$\begin{aligned}|R_n(t)|&\leqslant\sum_{k=1}^{n}\sum_{s=2}^{\infty}\frac{1}{2}|f_{nk}(t)-1|^s=\frac{1}{2}\sum_{k=1}^{n}\frac{|f_{nk}(t)-1|^2}{1-|f_{nk}(t)-1|}\\ &\leqslant\sum_{k=1}^{n}|f_{nk}(t)-1|^2\leqslant\max_{1\leqslant k\leqslant n}|f_{nk}(t)-1|\sum_{k=1}^{n}|f_{nk}(t)-1|;\end{aligned}$$

但由 (30) 中第一个不等式及 (24)

$$\sum_{k=1}^{n}|f_{nk}(t)-1| \leqslant \frac{t^2}{2}\sum_{k=1}^{n}\int_{-\infty}^{\infty}x^2dF_{nk}(x)=\frac{t^2}{2}.$$

故

$$|R_n(t) \leqslant \frac{t^2}{2}\max_{1\leqslant k\leqslant n}|f_{nk}(t)-1|.$$

由 (31) 可见当 $n\to\infty$ 时，关于任意有穷区间 $[-T,T]$ 中的 t 均匀地有

$$R_n(t)\to 0. \tag{34}$$

(乙) 令

$$\rho_n(t)=\frac{t^2}{2}+\sum_{k=1}^{n}\int_{-\infty}^{\infty}(e^{itx}-1-itx)dF_{nk}(x).$$

由 (29) 得

$$\sum_{k=1}^{n}(f_{nk}(t)-1)=-\frac{t^2}{2}+\rho_n(t). \tag{35}$$

如果能够证明：对任意有穷区间 $[-T,T]$ 中的 t, 均匀地有

$$\lim_{n\to\infty}\rho_n(t)=0. \tag{36}$$

那么以 (35) 代入 (28) 并联合 (甲) 中的结论即得证 (27), 而且 (27) 中的收敛对任意有穷区间内的 t 均匀，从而定理得以完全证明.

今证 (36), 由 (24)

$$\frac{t^2}{2}=-\sum_{k=1}^{n}\int_{-\infty}^{\infty}\frac{(itx)^2}{2}dF_{nk}(x).$$

对任意 $\varepsilon>0$,

$$\begin{aligned}\rho_n(t)=&\sum_{k=1}^{n}\int_{|x|\leqslant\varepsilon}\left[e^{itx}-1-itx-\frac{(itx)^2}{2}\right]dF_{nk}(x)\\&+\sum_{k=1}^{n}\int_{|x|>\varepsilon}\left[e^{itx}-1-itx+\frac{t^2x^2}{2}\right]dF_{nk}(x).\end{aligned}$$

由 (19)(18) 得

$$\begin{aligned}|\rho_n(t)|&\leqslant\frac{|t|^3}{6}\sum_{k=1}^{n}\int_{|x|\leqslant\varepsilon}|x|^3dF_{nk}(x)+t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2dF_{nk}(x)\\&\leqslant\frac{|t|^3}{6}\varepsilon\sum_{k=1}^{n}\int_{|x|\leqslant\varepsilon}x^2dF_{nk}(x)+t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2dF_{nk}(x).\end{aligned}$$

由 (24) 可见：对 $|t|\leqslant T$, 有

$$|\rho_n(t)|\leqslant\frac{|T|^3}{6}\varepsilon+T^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2dF_{nk}(x). \tag{37}$$

对任意 $\eta > 0$, 可选 $\varepsilon > 0$ 使

$$\frac{|T|^3}{6}\varepsilon < \frac{\eta}{2}.$$

又由 (25), 存在正整数 $N = N(T, \eta, \varepsilon)$, 使对此 ε 及 $n \geqslant N$, 有

$$\sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 \, dF_{nk}(x) < \frac{\eta}{2T^2} \tag{38}$$

于是当 $n \geqslant N$ 时，对一切 $t \in [-T, T]$, 有

$$|\rho_n(t)| < \eta. \quad \square$$

系 2 (李雅普诺夫 [Ляпунов]) 如对独立随机变数列 $\{\xi_k\}$, 存在常数 $\delta > 0$, 使当 $n \to \infty$ 时有

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|\xi_k - a_k|^{2+\delta} \to 0. \tag{39}$$

则 (4) 对 x 均匀地成立.

证 只要验证麟德伯尔格条件满足，由 (39)

$$\begin{aligned}
&\frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 dF_k(x) \\
&\leqslant \frac{1}{B_n^2 (\tau B_n)^\delta} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} |x-a_k|^{2+\delta} dF_k(x) \\
&\leqslant \frac{1}{\tau^\delta} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|\xi_k - a_k|^{2+\delta} \to 0, \ (n \to \infty). \quad \square
\end{aligned}$$

(四) 若干注释. 简单叙述中心极限定理的发展是有益的，从这里可以看出前人的辛勤劳动. 这一重要定理的研究开始于 18 世纪，首先是隶莫佛 (de Moivre 1667–1754) 对贝努利试验发现此定理 (见系 1), 拉普拉斯 (Laplace 1748–1827) 用了将近 20 年的时间来研究 $\zeta_n = \frac{1}{B_n} \sum\limits_{k=1}^{n} (\xi_k - a_k)$ 的极限分布，他就贝努利试验情形改进了隶莫佛的证明，并对相当一般的情形指出了极限分布是正态分布，然而他的证明以近代的眼光来看是不够严格的. 拉普拉斯的结果，首先由李雅普诺夫 (Ляпунов) 于 1901 年给出严格证明 (见系 2). "中心极限定理" 的名称由卜里耶于 1920 年提出. 自 1901 年起，许多人在这问题上做过工作，主要目标是研究使中心极限定理成立的最广泛的条件，直到 1922 年才有了显然的进展 — 麟德伯尔格提出了他的著名的条件. 人们真正认识到这条件的重要性是由于费勒 (Feller), 他于 1935 年发现：在独立随机变数列情况，这条件不仅是中心极限定理成立的充分条件，甚至在一定条件下还是必要的 (见下节系 1).

麟德伯尔格条件的价值在于它的广泛性，但应用时却不容易，一是计算复杂，二是在许多实际问题中，$F_k(x)$ 往往事先未给出，因而，上面几个系起着重要的作用.

定理 2 的一个直观 (然而不严格) 的解释如下：一个现实的量如果是由大量独立的而且均匀小的变量相加而成，那么它的分布近似于正态分布，这揭示了正态分布的重要性.

因为现实中许多变量都具有上述性质，例如，成年人身体的高度是受许多因素(先天的，后天的)影响的总结果，因而一般认为身高是有近似正态分布的随机变数. 同样道理，人体的重量，砖的抗压强度等等也如此.

大数定理与中心极限定理的关系如何？当 $\{\xi_k\}$ 独立，同分布，方差有穷且大于 0 时，容易讨论，因为这时二定理都成立. 大数定理断定：对任意 $\varepsilon>0$,

$$\lim_{n\to\infty}P\left(\left|\frac{\sum\limits_{k=1}^{n}(\xi_k-E\xi_k)}{n}\right|\leqslant\varepsilon\right)=1.$$

然而括号中事件的概率有多大？此定理未回答，但中心极限定理却给出一近似解答：

$$P\left(\left|\frac{\sum\limits_{k=1}^{n}(\xi_k-E\xi_k)}{n}\right|\leqslant\varepsilon\right)=P\left(\left|\frac{\sum\limits_{k=1}^{n}(\xi_k-E\xi_k)}{\sigma\sqrt{n}}\right|\leqslant\frac{\varepsilon\sqrt{n}}{\sigma}\right)$$
$$\approx\frac{1}{\sqrt{2\pi}}\int_{|x|\leqslant\frac{\varepsilon\sqrt{n}}{\sigma}}e^{-x^2/2}\,dx,$$

其中 $\sigma^2=D\xi_k$. 因而在所假定的条件下，中心极限定理比大数定理更精确.

*§3.5 中心极限定理(续)

(一) 麟德伯尔格条件的必要性. 设 $\{\xi_k\}$ 为独立随机变数列，ξ_k 的数学期望与方差为

$$a_k=E\xi_k,\ \ 0<b_k^2=D\xi_k<\infty\ (k=1,2,\cdots),\tag{1}$$

令 $B_n^2=\sum\limits_{k=1}^{n}b_k^2$, 上节中已看到，为了使中心极限定理成立，即为了对 $x\in R_1$ 均匀地有

$$\lim_{n\to\infty}P\left(\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k-a_k)\leqslant x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-z^2/2}\,dz.\tag{2}$$

充分条件是麟德伯尔格条件：对任意 $\tau>0$

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_k(x)=0.\tag{3}$$

现在来研究这条件的必要性，从而使中心极限定理取完善的形式.

引理 1 如麟德伯尔格条件满足，则

$$\lim_{n\to\infty}B_n=\infty,\tag{4}$$

$$\lim_{n\to\infty}\frac{b_n}{B_n}=0.\tag{5}$$

证 如 (4) 不成立，则必有常数 B, 使 $B_n\leqslant B$. 因 $b_1^2=\int_{-\infty}^{\infty}(x-a_1)^2\,dF_1(x)$, 故存在 $\varepsilon>0$, 使

$$\int_{|x-a_1|>\tau B}(x-a_1)^2\,dF_1(x)>\frac{b_1^2}{2}.$$

从而

$$\frac{1}{B_n}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_k(x)\geqslant\frac{1}{B}\int_{|x-a_1|>\tau B}(x-a_1)^2\,dF_1(x)>\frac{b_1^2}{2B}>0,$$

这与 (3) 矛盾而得证 (4).

其次，任取 τ, 使 $\frac{1}{2}>\tau>0$, 当 n 充分大时，由 (3)

$$\int_{|x-a_n|>\tau B_n}(x-a_n)^2\,dF_n(x)\leqslant\frac{\tau}{2}B_n^2.$$

故

$$b_n^2=\int_{-\infty}^{\infty}(x-a_n)^2\,dF_n(x)=\int_{|x-a_n|\leqslant\tau B_n}(x-a_n)^2\,dF_n(x)$$
$$+\int_{|x-a_n|>\tau B_n}(x-a_n)^2\,dF_n(x)\leqslant\tau^2B_n^2+\frac{\tau}{2}B_n^2<\tau B_n^2. \tag{6}$$

于是 $\frac{b_n^2}{B_n^2}<\tau$. □

引理 2 (4)(5) 二式成立的充要条件是

$$\lim_{n\to\infty}\max_{k\leqslant n}\frac{b_k}{B_n}=0. \tag{7}$$

证 设 (7) 成立，由 $\frac{b_n}{B_n}\leqslant\max\limits_{k\leqslant n}\frac{b_k}{B_n}$ 得 (5); 如说 $B_n\to B(<\infty)$, 则因 $\max\limits_{k\leqslant n}\frac{b_k}{B_n}\geqslant\frac{b_1}{B_n}$ 而与 (7) 矛盾.

设 (4)(5) 成立. 对任意 $\varepsilon>0$, 取正整数 M, 使 $\frac{b_n}{B_n}<\varepsilon$ 对一切 $n\geqslant M$ 成立；由 (4) 可取正整数 $N\geqslant M$, 使 $\max\limits_{k\leqslant M}\frac{b_k}{B_N}<\varepsilon$, 于是对一切 $n\geqslant N$ 有

$$\max_{k\leqslant M}\frac{b_k}{B_n}\leqslant\max_{k\leqslant M}\frac{b_k}{B_N}<\varepsilon,$$

并且对 k: $M<k\leqslant n$ 有

$$\frac{b_k}{B_n}\leqslant\frac{b_k}{B_k}<\varepsilon.\quad\square$$

定理 1 (**麟德伯尔格 - 费勒**) 对独立随机变数列 $\{\xi_k\}$, 中心极限定理及 (4)(5) 成立的充要条件是麟德伯尔格条件满足.

证 充分性由上节定理 2 及本节引理 1 推出. 下证必要性.

仍以 $f_{nk}(t)$ 及 $F_{nk}(x)$ 分别表 $\xi_{nk}=\frac{\xi_k-a_k}{B_n}$ 的特征函数与分布函数，因中心极限定理成立，故

$$\prod_{k=1}^{n}f_{nk}(t)\to e^{-\frac{t^2}{2}},\ (n\to\infty) \tag{8}$$

以 θ 表某复常数， $|\theta|\leqslant 1$, 它的精确值在每次出现时可不同，但不必明确指出. 我们有

$$f_{nk}(t)=\int_{-\infty}^{\infty}(1+itx+\frac{1}{2}\theta t^2x^2)\,dF_{nk}(x)=1+\frac{1}{2}\theta t^2\frac{b_k^2}{B_n^2}. \tag{9}$$

由引理 2 得

$$\max_{k\leqslant n}|f_{nk}(t)-1|\to 0\ (n\to\infty). \tag{10}$$

由此及 (9) 知

$$\begin{aligned}\sum_{k\leqslant n}|f_{nk}(t)-1|^2 &\leqslant \max_{k\leqslant n}|f_{nk}(t)-1|\sum_{k=1}^{n}|f_{nk}(t)-1|\\ &=\max_{k\leqslant n}|f_{nk}(t)-1|\frac{1}{2}|\theta|t^2\to 0\ (n\to\infty).\end{aligned} \tag{11}$$

故当 n 充分大时， $\log f_{nk}(t)$ 存在，而 (8) 式化为

$$\sum_{k=1}^{n}\log f_{nk}(t)\to -\frac{t^2}{2}\ (n\to\infty). \tag{12}$$

利用 $\log z=z-1+\theta|z-1|^2$, 得

$$\log f_{nk}(t)=f_{nk}(t)-1+\theta|f_{nk}(t)-1|^2.$$

代入 (12) 并注意 (11), 可见对任意 $t\in R_1$,

$$\left|\frac{t^2}{2}-\sum_{k=1}^{n}\{1-f_{nk}(t)\}\right|\to 0\ (n\to\infty).$$

取此式的实部，对任意 $\tau>0$, 有

$$\frac{t^2}{2}-\sum_{k=1}^{n}\int_{|x|\leqslant\tau}(1-\cos tx)\,dF_{nk}(x)=\sum_{k=1}^{n}\int_{|x|>\tau}(1-\cos tx)\,dF_{nk}(x)+o(1). \tag{13}$$

由于 $1-\cos y\leqslant\frac{y^2}{2}$, $(y\in R_1)$ 及上节 (24),(13) 中左方积分

$$\begin{aligned}\sum_{k=1}^{n}\int_{|x|\leqslant\tau}(1-\cos tx)\,dF_{nk}(x) &\leqslant \frac{t^2}{2}\sum_{k=1}^{n}\int_{|x|\leqslant\tau}x^2\,dF_{nk}(x)\\ &=\frac{t^2}{2}\left[1-\sum_{k=1}^{n}\int_{|x|>\tau}x^2\,dF_{nk}(x)\right].\end{aligned}$$

(13) 中右方积分

$$\begin{aligned}\sum_{k=1}^{n}\int_{|x|>\tau}(1-\cos tx)\,dF_{nk}(x) &\leqslant 2\sum_{k=1}^{n}\int_{|x|>\tau}dF_{nk}(x)\\ &\leqslant\frac{2}{\tau^2}\sum_{k=1}^{n}\int_{|x|>\tau}x^2\,dF_{nk}(x)\leqslant\frac{2}{\tau^2}.\end{aligned}$$

故由 (13) 得

$$\frac{t^2}{2}\sum_{k=1}^{n}\int_{|x|>\tau}x^2\,dF_{nk}(x)\leqslant\frac{2}{\tau^2}+o(1),$$

或

$$0 \leqslant \sum_{k=1}^{n} \int_{|x|>\tau} x^2 \, dF_{nk}(x) \leqslant \frac{2}{t^2}\Big[\frac{2}{\tau^2} + o(1)\Big].$$

先令 $n \to \infty$, 再令 $t \to \infty$, 即得上节 (25), 亦即得证麟德伯尔格条件满足. □

系 1 对独立随机变数列 $\{\xi_k\}$, 如 (4)(5) 满足，则中心极限定理成立的充要条件是麟德伯乐格条件满足.

由定理 1 直接推出. □

(二) **例.**

例 1 设 $\{\xi_k\}$ 为独立随机变数列， ξ_k 有密度矩阵为

$$\begin{pmatrix} -k^{\alpha} & 0 & k^{\alpha} \\ \frac{1}{2k^{2\alpha}} & 1-\frac{1}{k^{2\alpha}} & \frac{1}{2k^{2\alpha}} \end{pmatrix}. \tag{14}$$

因而 $E\xi_k = 0$, $D\xi_k = 1$, $B_n^2 = \sum\limits_{k=1}^{n} D\xi_k = n$, 由此可见条件 (4)(5) 满足，由系 1, 这时中心极限定理成立的充要条件是：对任意 $\tau > 0$.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\substack{1\leqslant k\leqslant n \\ k^{\alpha}>\tau\sqrt{n}}} I(k) = 0, \tag{15}$$

其中 $I(k) \equiv 1$, 而求和对一切同时满足 $1 \leqslant k \leqslant n$, $k^{\alpha} > \tau\sqrt{n}$ 的 k 进行. 容易看出，条件 (15) 当 $\alpha < \frac{1}{2}$ 时满足，而当 $\alpha \geqslant \frac{1}{2}$ 不满足.

如不考虑 $\alpha = \frac{1}{2}$, 还可以用另一方法证明 $\alpha > \frac{1}{2}$ 时中心极限定理不成立. 由 (14) 可见

$$P\Big(\sum_{k=2}^{n} \xi_k = 0\Big) > \prod_{k=2}^{\infty}\Big(1 - \frac{1}{k^{2\alpha}}\Big) > 0,$$

故

$$\lim_{n\to\infty} P\left(\frac{\sum\limits_{k=1}^{n}\xi_k}{\sqrt{n}} = 0\right) = \lim_{n\to\infty} P\left(\frac{\sum\limits_{k=2}^{n}\xi_k}{\sqrt{n}} = 0\right) = \lim_{n\to\infty} P\Big(\sum_{k=2}^{n}\xi_k = 0\Big) > 0.$$

如说中心极限定理成立，左方极限应等于 0 而与上式矛盾.

例 2 麟德伯尔格条件不满足，中心极限定理仍可成立. 设 $\{\xi_k\}$ 为独立随机变数列, ξ_k 有正态分布为 $N(0, b_k)$, $b_k^2 > 0$, $\sum\limits_{k=1}^{\infty} b_k^2 < \infty$. 这时 (4) 不成立，因而由引理 1 麟德伯尔格条件不满足，但根据 $N(a_1, b_1) \times N(a_2, b_2) = N(a_1 + a_2, \sqrt{b_1^2 + b_2^2})$, 知 $\frac{\sum\limits_{k=1}^{n}\xi_k}{B_n}$ 有正态分布 $N(0,1)$, 故

$$P\left(\frac{\sum\limits_{k=1}^{n}\xi_k}{B_n} \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} \, dz.$$

有趣的是反面的结果也正确：设 $\{\xi_k\}$ 为独立随机变数列， $\sum\limits_{n=1}^{\infty} D\xi_n < \infty$, 如果中心极限定理成立，则每 ξ_n 有正态分布 (参看 [12] 第七章 §4 或 [14] 第六章， 3 段).

*§3.6 格子点分布与局部极限定理

(一) 问题的提出．在 §3.4 系 1 中，我们已经看到，如果 $\{\xi_k\}$ 是一列独立同分布的随机变数，ξ_k 的分布有密度为 $\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$, $0<p<1$, 令 $\eta_n=\sum\limits_{k=1}^{n}\xi_k$, 则对 $a,b(a<b)$ 均匀地有

$$\lim_{n\to\infty} P\left(a<\frac{\eta_n-np}{\sqrt{npq}}\leqslant b\right)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{z^2}{2}}\,dz. \tag{1}$$

如记 $P_n(k)=P(\eta_n=k)$, 则上式化为

$$\lim_{n\to\infty}\sum_{np+a\sqrt{npq}<k\leqslant np+b\sqrt{npq}} P_n(k)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{z^2}{2}}\,dz. \tag{2}$$

现在希望对固定的一个 k, 近似地计算 $P_n(k)$. 为此，在 (2) 中形式地令

$$a=\frac{k-1-np}{\sqrt{npq}},\quad b=\frac{k-np}{\sqrt{npq}}. \tag{3}$$

并用中值定理，得

$$P_n(k)=\frac{1}{\sqrt{2\pi}}\int_{\frac{k-1-np}{\sqrt{npq}}}^{\frac{k-np}{\sqrt{npq}}} e^{-\frac{z^2}{2}}\,dz\approx\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}e^{-(z_{nk})^2/2}, \tag{4}$$

其中 $z_{nk}=\frac{k-np}{\sqrt{npq}}$, "$\approx$" 表 "近似".

上述推导启发我们，(4) 式也许真是 $P_n(k)$ 的近似式，但要注意，不能认为 (4) 式已经严格证明了.

虽然如此，下面会严格证明 (4) 式的确成立.

回忆在 §2.4 中，我们用普阿松分布也给出了 $P_n(k)$ 的一个近似式.

为了证明以 (4) 为特例的一般结果，先引进格子点分布的概念.

(二) 格子点分布．称离散型分布 $F(A)$ $(A\in\mathcal{B}_1)$ 为格子点分布，如果它的密度矩阵可表为

$$\begin{pmatrix} a+kh \\ p_k \end{pmatrix};\quad k=\cdots,-2,-1,0,1,2,\cdots, \tag{5}$$

其中 a 及 $h>0$ 为常数，k 可以取任意整数为值，并称 h 为分布的步长.

二项分布，普阿松分布，几何分布等都是格子点分布.

试用特征函数来表达分布的格子点性质

引理 1 分布 $F(A)$, $(A\in\mathcal{B}_1)$ 为格子点分布的充要条件是：它的特征函数的绝对值在某 $t\neq 0$ 上等于 1.

证 此分布的特征函数为

$$f(t)=\sum_{-\infty}^{\infty}p_k e^{it(a+kh)}=e^{iat}\sum_{k=-\infty}^{\infty}p_k e^{itkh}, \tag{6}$$

$$\left|f\left(\frac{2\pi}{h}\right)\right|=\left|e^{2\pi i\frac{a}{h}}\sum_{-\infty}^{\infty}p_k e^{2\pi ik}\right|=\left|e^{2\pi i\frac{a}{h}}\right|=1. \tag{7}$$

反之，设分布 $F(A)$ $(A\in\mathcal{B}_1)$ 的分布函数为 $F(x)$, 特征函数 $f(t)$ 在某 $t_1\neq 0$ 上有 $|f(t_1)|=1$. 于是

$$\int_{-\infty}^{\infty}e^{it_1x}dF(x)=f(t_1)=e^{i\theta},$$

θ 为某常数，由上式得

$$\int_{-\infty}^{\infty}e^{i(t_1x-\theta)}\,dF(x)=1. \tag{8}$$

从而

$$\int_{-\infty}^{\infty}[1-\cos(t_1x-\theta)]\,dF(x)=0.$$

由于 $1-\cos(t_1x-\theta)$ 连续而且非负，为使上式成立，分布 $F(A)(A\in\mathcal{B}_1)$ 必须集中在集 $D=(x:\ \cos(t_1x-\theta)=1)$ 上，亦即集中在形如 $x_k=\frac{\theta}{t_1}+k\frac{2\pi}{t_1}$ 的点的集 $\{x_k\}$ 上，k 为整数，故此分布是格子点分布. □

称步长 h 为最大的，如果对任意实数 b 及 $h_1>h$, 此分布所集中的可列点集 Λ 不能表为 $\{b+kh_1\}$ 的形式. 以 B 表 Λ 中任意两点的差的集，则 h 为最大步长的充要条件是 B 中元除以 h 以后的最大公因子为 1.

例如，设分布集中在集 $\Lambda=\{4k\}$ 上，即 4 的倍数的点集上，那么，1,2,4 都是步长，而 4 是最大步长.

引理 2 设 h 是最大步长，则对任意 $\varepsilon>0$, 存在 $c>0$, 使

$$|f(t)|\leqslant e^{-c},\ \text{如}\ \varepsilon\leqslant|t|\leqslant\frac{2\pi}{h}-\varepsilon. \tag{9}$$

先证

$$|f(t)|<1,\ \text{如}\ 0<|t|<\frac{2\pi}{h}. \tag{10}$$

设若不然，存在 t_1, $0<|t_1|<\frac{2\pi}{h}$, 使 $|f(t_1)|=1$, 由引理 1 的证明可见 $\frac{2\pi}{|t_1|}$ 是步长；然而 $\frac{2\pi}{|t_1|}>h$ 这与 h 是最大步长矛盾. 根据 $f(x)$ 的连续性，(9) 式自 (7) 及 (10) 推出. □

(三) **局部极限定理** 设 $\{\xi_k\}$ 为独立随机变数列，有相同的格子点分布，其密度为 (5). 考虑部分和

$$\eta_n=\sum_{k=1}^{n}\xi_k.$$

显然，η_n 的可能值 (即以正概率能取的值) 可表为 $na+kh$ 的形式. 令

$$P_n(k)=P(\eta_n=na+kh), \tag{11}$$

$$z_{n,k}=\frac{an+kh-A_n}{B_n}, \tag{12}$$

其中 $A_n=E\eta_n=nE\xi_1=nm$, $B_n^2=D\eta_n=nD\xi_1=n\sigma^2>0(E\xi_1=m, D\xi_1=\sigma^2)$.

定理 1 (格涅坚科 [Гнеденко]) 设 $\{\xi_k\}$ 为独立随机变数列，有相同的格子点分布及有穷的数学期望与方差，为使对 $k(-\infty<k<\infty)$ 均匀地有

$$\lim_{n\to\infty}\left(\frac{B_n}{h}P_n(k)-\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{n,k}^2}{2}}\right)=0, \tag{13}$$

充要条件是步长为最大的.

证 充分：由 (6) 及独立同分布性知 η_n 的特征函数为

$$f^n(t)=\left[e^{iat}\sum_{k=-\infty}^{\infty}p_k e^{itkh}\right]^n=e^{iant}\sum_{k=-\infty}^{\infty}P_n(k)e^{itkh}.$$

以 $e^{-iant-itkh}$ 乘两边并自 $-\frac{\pi}{h}$ 积分到 $\frac{\pi}{h}$, 得

$$\frac{2\pi}{h}P_n(k)=\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}f^n(t)e^{-iant-itkh}\,dt.$$

简记 (12) 中的 $z_{n,k}$ 为 z, 得

$$\begin{aligned}hk&=B_nz+A_n-an,\\ \frac{2\pi}{h}P_n(k)&=\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}g^n(t)e^{-itzB_n}\,dt,\end{aligned}$$

其中 $g(t)=e^{-\frac{itA_n}{n}}f(t)$, 再令 $x=tB_n$, 有

$$\frac{2\pi B_n}{h}P_n(k)=\int_{-\frac{\pi B_n}{h}}^{\frac{\pi B_n}{h}}e^{-izx}g^n\left(\frac{x}{B_n}\right)dx. \tag{14}$$

现在来估计

$$R_n=2\pi\left[\frac{B_n}{h}P_n(k)-\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\right]. \tag{15}$$

为此，以 (14) 及 $e^{-\frac{z^2}{2}}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-izx-\frac{x^2}{2}}\,dx$ (见 §2.11 例 3) 代入上式，并把 R_n 拆为

$$R_n=\sum_{j=1}^{4}I_j, \tag{16}$$

其中

$$\begin{aligned}I_1&=\int_{-A}^{A}e^{-izx}\left[g^n\left(\frac{x}{B_n}\right)-e^{-\frac{x^2}{2}}\right]dx,\quad I_2=-\int_{|x|>A}e^{-izx-\frac{x^2}{2}}\,dx,\\ I_3&=\int_{\varepsilon B_n\leqslant|x|\leqslant\frac{\pi B_n}{h}}e^{-izx}g^n\left(\frac{x}{B_n}\right)dx,\quad I_4=\int_{A\leqslant|x|<\varepsilon B_n}e^{-izx}g^n\left(\frac{x}{B_n}\right)dx,\end{aligned}$$

这里 $A>0,\varepsilon>0$ 是常数，它们留待下面选定. 现在来分别估计 I_j 的值，由 §3.4 定理 1, 对任意有穷区间内的 t, 均匀地有 $\lim\limits_{n\to\infty}g^n\left(\frac{t}{B_n}\right)=e^{-t^2/2}$, 从而不论 $A>0$ 如何，有

$$\lim_{n\to\infty}I_1=0. \tag{17}$$

选 A 充分大后，可使 $|I_2|$ 充分小，这是因为

$$|I_2|\leqslant\int_{|x|>A}e^{-\frac{x^2}{2}}\,dx\leqslant\frac{2}{A}\int_{A}^{\infty}xe^{-\frac{x^2}{2}}\,dx=\frac{2}{A}e^{-\frac{A^2}{2}}. \tag{18}$$

由 (9) 得

$$|I_3| \leqslant \int_{\varepsilon B_n \leqslant |x| \leqslant \frac{\pi B_n}{h}} \left|g\left(\frac{x}{B_n}\right)\right|^n dx \leqslant 2e^{-nc} B_n\left(\frac{\pi}{h} - \varepsilon\right). \tag{19}$$

最后，由方差有穷的假定知 $g(t)$ 有二级导数，由 §2.11, 在 $t=0$ 附近有展式

$$g(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

当 $\varepsilon > 0$ 充分小，对 t 有

$$|g(t)| < 1 - \frac{\sigma^2 t^2}{4} < e^{-\frac{\sigma^2 t^2}{4}} \quad (\text{如}|t| < \varepsilon),$$

$$\left|g\left(\frac{x}{B_n}\right)\right|^n < e^{-\frac{n\sigma^2 t^2}{4B_n^2}} = e^{-\frac{t^2}{4}} \quad (\text{如}|x| < \varepsilon B_n).$$

故

$$|I_4| \leqslant 2\int_A^{\varepsilon B_n} e^{-\frac{t^2}{4}}\, dt < 2\int_A^{\infty} e^{-\frac{t^2}{4}}\, dt. \tag{20}$$

因而当 A 充分大后， $|I_4|$ 可充分小.

必要：设 h 是步长但非最大的，则 B 中元除以 h 以后的最大公因子 $d > 1$. 而且 η_n 的可能值可表为 $na + kdh$ 之形， k 为整数，如能选整数 $l = l(n)$, 使 $ld+1$ 满足条件

$$z_{n,ld+1} = \frac{an + (ld+1)h - A_n}{B_n} \to 0, \quad (n \to \infty) \tag{21}$$

那么，一方面由于数 $na + (ld+1)h$, 不呈 $na + kdh$ 之形，故 $P_n(ld+1) = 0$; 另方面，由 (21) 又有

$$\lim_{n\to\infty} e^{-\frac{z_{n,ld+1}^2}{2}} = 1.$$

于是 (13) 不成立而必要性得以证明.

现在来选 l, 首先注意 (5) 中的 a 不是唯一的，例如可以用 $a+h$ 代替 a, 故不妨设 $a \neq m = E\xi_1$, 回忆 $A_n = nm$, $B_n^2 = n\sigma^2$, 可见选

$$l = \left[\frac{(m-a)^n}{dh}\right] \tag{22}$$

就可使 (21) 成立，这里 $[b]$ 表不大于 b 的最小整数. □

由于二项分布的最大步长为 1, 故由定理 1 直接得

系 1 (隶莫佛 - 拉普拉斯) 设 $\{\xi_k\}$ 为独立同分布的随机变数列， ξ_k 的分布密度为 $\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$, $0 < p < 1$, 令 $P_n(k) = P\left(\sum\limits_{j=1}^{n} \xi_j = k\right)$, 则关于非负整数 k 均匀地有

$$\lim_{n\to\infty}\left(\sqrt{npq}\cdot P_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{nk}^2}{2}}\right) = 0, \tag{23}$$

其中 $z_{nk} = \frac{k-np}{\sqrt{npq}}$.

关于局部极限定理的更多的研究见 [11].

*§3.7 若 干 补 充

(一) 特征函数的充要条件. 函数 $f(t)(t \in R_1)$ 是特征函数的充要条件是什么？前面已看到，一些简单的必要条件是：$f(x)$ 均匀连续，$|f(t)| \leqslant f(0) = 1$, $f(-t) = \overline{f(t)}$, 非负定等等.

关于充分条件与充要条件方面也有不少结果 [1], 我们只叙述其中重要的一个.

波赫纳尔 - 辛软 (Bochner-Хинчин) **定理**. 函数 $f(t)(t \in R_1)$ 是特征函数的充要条件是：$f(t)$ 非负定，连续而且 $f(0) = 1$.

在证明的过程中，附带可以证明一个与上定理类似的赫尔格洛茨 (Herglotz) 定理. 称复数列 $\{c_n\}$ $n = 0, \pm 1, \pm 2, \cdots$ 为非负定的，如

$$\sum_{k=1}^{n}\sum_{j=1}^{n} c_{k-j}\xi_k\bar{\xi}_j \geqslant 0.$$

对任意正整数 n 及任意复数 $\xi_1, \cdots, \xi_n$ 成立.

赫尔格洛茨定理. 数列 $\{c_n\}$ 可表为

$$c_n = \int_{-\pi}^{\pi} e^{inx} dG(x) \quad (n = 0, \pm 1, \pm 2, \cdots)$$

的充要条件是它为非负定的，这里 $G(x)$ 是 $[-\pi, \pi]$ 上有界，单调不减，右连续函数.

这定理的必要性部分证明完全与 §2.11(7) 的证明类似.

波赫纳尔 - 辛钦定理的证 由上述只要证充分性.

任取正整数 N, 对 $\frac{k}{n}$ 及 $e^{-ikx}(k = 0, 1, \cdots, N-1)$, 由非负定性有

$$\mathcal{F}_N^{(n)}(x) = \frac{1}{N}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1} f\left(\frac{k-j}{n}\right)e^{-i(k-j)x} \geqslant 0.$$

不难算得，在这和中使 $k - j = r$ 的项共有 $N - |r|$ 个，r 可自 $-N+1$ 变到 $N-1$. 因此

$$\mathcal{F}_N^{(n)}(x) = \sum_{r=-N}^{N}\left(1 - \frac{|r|}{N}\right) f\left(\frac{r}{n}\right) e^{-irx},$$

$$\int_{-\pi}^{\pi} e^{isx}\mathcal{F}_N^{(n)}(x)\, dx = \sum_{r=-N}^{N}\left(1 - \frac{|r|}{N}\right) f\left(\frac{r}{n}\right)\int_{-\pi}^{\pi} e^{i(s-r)x}\, dx.$$

利用 $\int_{-\pi}^{\pi} e^{iux}\, dx = 0$ 或 2π, 视 $u \neq 0$ 或 $u = 0$, 而定，得

$$\left(1 - \frac{|s|}{N}\right) f\left(\frac{s}{n}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{isx}\mathcal{F}_N^{(n)}(x)\, dx = \int_{-\pi}^{\pi} e^{isx} dF_N^{(n)}(x), \tag{1}$$

其中 $F_N^{(n)}(x) = \frac{1}{2\pi}\int_{-\pi}^{x}\mathcal{F}_N^{(n)}(t)\, dt$ 是 $[-\pi, \pi]$ 上有界单调不减函数，全变差为

$$F_N^{(n)}(\pi) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{F}_N^{(n)}(t)\, dt = f(0) = 1.$$

[1] 参看 [5].

补定义 $F_N^{(n)}(x)=0,\ x<-\pi; F_N^{(n)}(x)=1,\ x>\pi$, 则 $F_N^{(n)}(x)$ 是一分布函数.

由海来第一定理 (§3.2), 存在序列 $N_k, N_k\to\infty(k\to\infty)$, 使 $\{F_N^{(n)}(x)\}$ 弱收敛于某单调不减, 右连续函数 $F^{(n)}(x)$. 又因对任意 N 及 $\varepsilon>0$ 有

$$F_N^{(n)}(-\pi-\varepsilon)=0,\quad F_N^{(n)}(\pi+\varepsilon)=1.$$

故 $F^{(n)}(-\pi-\varepsilon)=0,\ F^{(n)}(\pi+\varepsilon)=1$ 而 $F^{(n)}(x)$ 还是一分布函数.

由 §3.2 定理 2,

$$\lim_{k\to\infty}\int_{-\pi}^{\pi}e^{isx}dF_{N_k}^{(n)}(x)=\int_{-\pi}^{\pi}e^{isx}dF^{(n)}(x),$$

因而根据 (1) 并注意 $\frac{|s|}{N_k}\to 0(k\to\infty)$, 得

$$f\left(\frac{s}{n}\right)=\int_{-\pi}^{\pi}e^{isx}dF^{(n)}(x) \tag{2}$$

(至此已证完赫尔格洛茨定理).

考虑分布函数 $F_n(x)=F^{(n)}\left(\frac{x}{n}\right)$ 及特征函数

$$f_n(t)=\int_{-n\pi}^{n\pi}e^{itx}dF^{(n)}(x).$$

显见对任一整数 k 有

$$f_n\left(\frac{k}{n}\right)=f\left(\frac{k}{n}\right). \tag{3}$$

对任意 $t\in R_1$, 总可取整数 k(k 依赖于 n,t), 使

$$0\leqslant t-\frac{k}{n}<\frac{1}{n}.$$

由于 $f(t)$ 连续, 故

$$f(t)=\lim_{n\to\infty}f\left(\frac{k}{n}\right)=\lim_{n\to\infty}f_n\left(\frac{k}{n}\right). \tag{4}$$

如能证明对一切 $t\in R_1$, 有

$$f(t)=\lim_{n\to\infty}f_n(t) \tag{5}$$

则由 §3.2 定理 3 即知 $f(t)$ 是特征函数.

为证 (5), 由 (3)(4) 得

$$\begin{aligned}\lim_{n\to\infty}f_n(t)&=\lim_{n\to\infty}\left\{\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]+f_n\left(\frac{k}{n}\right)\right\}\\&=f(t)+\lim_{n\to\infty}\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right].\end{aligned} \tag{6}$$

令 $\theta=t-\frac{k}{n}$, 则 $0\leqslant\theta<\frac{1}{n}$. 根据 $f_n(t)$ 的定义, 得

$$\begin{aligned}\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|&=\left|\int_{-n\pi}^{n\pi}e^{i\frac{k}{n}x}(e^{i\theta x}-1)\,dF_n(x)\right|\\&\leqslant\int_{-n\pi}^{n\pi}|e^{i\theta x}-1|\,dF_n(x).\end{aligned} \tag{7}$$

由布尼亚科夫斯基 (Буняковский) 不等式

$$\begin{aligned}\int_{-n\pi}^{n\pi}|e^{i\theta x}-1|\,dF_n(x)&\leqslant\left[\int_{-n\pi}^{n\pi}|e^{i\theta x}-1|^2\,dF_n(x)\right]^{\frac{1}{2}}\\&=\left[\int_{-n\pi}^{n\pi}2(1-\cos\,\theta x)\,dF_n(x)\right]^{\frac{1}{2}}=[2(1-\operatorname{Re}f_n(\theta))]^{\frac{1}{2}},\end{aligned}\tag{8}$$

这里 $\operatorname{Re}f_n(\theta)$ 为 $f_n(\theta)$ 的实数部分. 因为对 $0\leqslant\alpha<1$ 及 $-\pi\leqslant z\leqslant\pi$ 有 $\cos\,z\leqslant\cos\,\alpha z$, 故

$$\begin{aligned}1-\operatorname{Re}f_n(\theta)&=\int_{-n\pi}^{n\pi}(1-\cos\,\theta x)\,dF_n(x)=\int_{-\pi}^{\pi}(1-\cos\,\theta nz)\,dF_n(nz)\\&\leqslant\int_{-\pi}^{\pi}(1-\cos\,z)\,dF_n(nz)=\int_{-\pi}^{\pi}(1-\cos\,z)\,dF^{(n)}(z)\\&=1-\operatorname{Re}\int_{-\pi}^{\pi}e^{iz}\,dF^{(n)}(z).\end{aligned}$$

由 (2)(取 $s=1$) 得

$$1-\operatorname{Re}f_n(\theta)\leqslant1-\operatorname{Re}f\left(\frac{1}{n}\right).\tag{9}$$

合并 (7)(8)(9), 即得

$$\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|\leqslant\left[2\left(1-\operatorname{Re}f\left(\frac{1}{n}\right)\right)\right]^{\frac{1}{2}}.$$

注意 $f(t)$ 连续而且 $f(0)=1$, 可见 $\lim\limits_{n\to\infty}\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|=0$. 于是由 (6) 知 (5) 成立

(二) **海来第二定理**. 关于分布函数列的弱收敛, 还有下列定理.

定理 1 分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充要条件是: 对任一有界连续函数 $f(x)$, 有

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}f(x)\,dF_n(x)=\int_{-\infty}^{\infty}f(x)\,dF(x).\tag{10}$$

证 必要: 设 $F_n(x)\to F(x)(\mathrm{w})$, 为了证明 (10), 只需重复 §3.2 定理 2 的证明, 作一些形式上的修改即可 [1].

充分: 设 (10) 成立, 任取 $F(x)$ 的连续点 b, 对任意 $\varepsilon>0$, 存在 $\delta>0$, 使

$$0\leqslant F(b+\delta)-F(b)<\varepsilon,\tag{11}$$

$$0\leqslant F(b)-F(b-\delta)<\varepsilon,\tag{12}$$

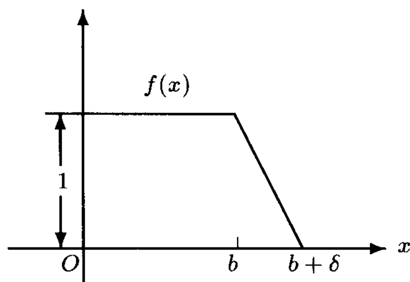


图 1

考虑有界连续函数 $f(x)$: 当 $x\leqslant b$ 时 $f(x)=1$; 当 $x\geqslant b+\delta$ 时, $f(x)=0$; 当 $x\in(b,b+\delta)$ 时, $f(x)$ 的值在 $[0,1]$ 之中, (例如, 图 1 中的 $f(x)$). 由 (10)(11) 得

[1] 参看 [8]§17 或 [9]§36.

$$
\begin{aligned}
\overline{\lim_{n\to\infty}} F_n(b) &\leqslant \overline{\lim_{n\to\infty}} \int_{-\infty}^{\infty} f(x)\,dF_n(x) \\
&= \int_{-\infty}^{\infty} f(x)\,dF(x) \leqslant F(b+\delta) \leqslant F(b)+\varepsilon.
\end{aligned} \tag{13}
$$

类似地，考虑有界连续函数 $g(x)$: 它在 $(-\infty, b-\delta]$ 及 $[b,\infty)$ 中分别取值 1 及 0, 在 $(b-\delta, b)$ 中取 0 与 1 之间的值，则

$$
\begin{aligned}
\underline{\lim_{n\to\infty}} F_n(b) &\geqslant \underline{\lim_{n\to\infty}} \int_{-\infty}^{\infty} g(x)\,dF_n(x) = \int_{-\infty}^{\infty} g(x)\,dF(x) \\
&\geqslant F(b-\delta) > F(b)-\varepsilon.
\end{aligned} \tag{14}
$$

综合 (13)(14) 得

$$
F(b)-\varepsilon < \underline{\lim_{n\to\infty}} F_n(b) \leqslant \overline{\lim_{n\to\infty}} F_n(b) < F(b)+\varepsilon,
$$

既然 $\varepsilon > 0$ 任意，故得证

$$
\lim_{n\to\infty} F_n(b) = F(b). \quad \square
$$

定理 1 的必要部分通常称为海来第二定理.

(三) **重对数定理**. 在极限定理中，除了中心极限定理与 (强) 大数定理外，还有一个著名的极限定理：重对数定理，这定理最初由辛钦于 1926 年对贝努里型试验证明，后来被柯尔莫哥洛夫等人推广.

我们只就贝努里型试验来说明这定理的实质.

以 η_n 表前 n 次试验中 A 出现的次数. $0 < p = P(A) < 1$, 强大数定理断定：

$$
\lim_{n\to\infty} \frac{\eta_n}{n} = p. \quad \text{(a.s.)}
$$

出此可见，以概率 1, 对任意 $\varepsilon > 0$,

$$
\left|\frac{\eta_n}{n} - p\right| < \varepsilon,
$$

对一切充分大的 n 都成立. 如果用 y 轴来表 $\eta_n - np$, 上式表示：以概率 1, $\eta_n - np$ 界于二直线 $y = \varepsilon n$ 及 $y = -\varepsilon n$ 之中，除对有穷多个 n 以外；就是说，以概率 1,$\eta_n - np$ 只能走出此二直线以外有穷次. 然而我们进一步问：这样的界线是否太宽？能否指出最可能精确的界线？

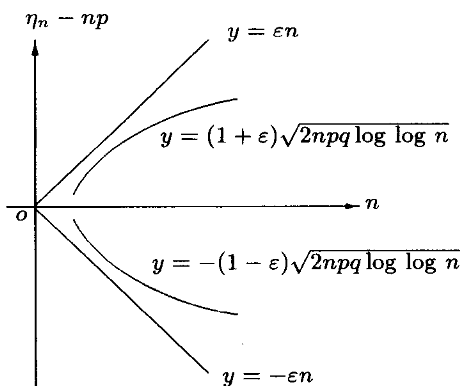


图 2

结果发现：可以把这两条线改为

$$
y = (1+\varepsilon)\sqrt{2npq\log\log n}, \quad y = -(1+\varepsilon)\sqrt{2npq\log\log n},
$$

这里 $q = 1-p$ 而 $\varepsilon > 0$ 任意. 对于这两条新线所围成区域，以概率 1,$\eta_n - np$ 也至多只能走出有穷次；而且还证明了：以概率 1,$\eta_n - np$ 会走出下二线

$$
y = (1-\varepsilon)\sqrt{2npq\log\log n}, \quad y = -(1-\varepsilon)\sqrt{2npq\log\log n},
$$

无穷多次.

这就是下定理的基本含义.

重对数定理. (辛钦) 对贝努里试验 $\widetilde{E}=E^{\infty}$, 以 η_n 表 A 在前 n 次试验中出现次数, A 在每次试验中出现的概率为 $p, 0<p<1, q=1-p$ 则

$$\varlimsup_{n\to\infty}\frac{\eta_n-np}{\sqrt{2npq\log\log n}}=1,\quad \text{(a.s.)}$$

$$\varliminf_{n\to\infty}\frac{\eta_n-np}{\sqrt{2npq\log\log n}}=1,\quad \text{(a.s.)}$$

证明从略 (见 [15] VIII, 5, 192 页这定理的推广见 [2] 260 页).

习　题

1. 设 $\{A_n\}$ 是独立事件列 (就是说, $\{A_n\}$ 中任意有限多个事件都相互独立), 而且 $\sum\limits_{n=1}^{\infty}P(A_n)=\infty$, 试证

$$P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right)=1.$$

(注: 本题连同 §3.1 引理 1 合称为 Borel-Cantelli 引理).

2. 设 $f(x)$ 是 $(0,\infty)$ 上的连续, 严格单调上升函数, $f(0)=0, \sup\limits_{x\geqslant 0}f(x)<\infty$.

证明: $\xi_n\to 0,$(P) 的充要条件是 $\lim\limits_{n\to\infty}E[f(|\xi_n|)]=0$.

3. 求证: 设 $\varphi(t)$ 是特征函数, 如果 $\psi(t)$ 满足: 对于某一序列 $\{h_n\}$ $(h_n\to\infty)$, $\varphi(t)\psi(h_nt)=f_n(t)$ 也是特征函数, 则 $\psi(t)$ 也是特征函数.

4. 求证: 如果 $f(t)$ 是特征函数, 则 $\varphi(t)=e^{f(t)-1}$ 也是特征函数.

5. 设 ξ_n 的分布是参数为 n 的 Poisson 分布, 求证: 当 $n\to\infty$ 时, $\frac{\xi_n-n}{\sqrt{n}}$ 的分布弱收敛于 $N(0,1)$.

6. 将 n 个球投入 N 个匣子, 球不可辨, 匣可辨. 设某个指定的匣子中恰有 k 个球的概率是 q_k, 求证: 当 $n\to\infty, N\to\infty, \frac{n}{N}\to\lambda$ 时,

$$q_k\longrightarrow\frac{\lambda^k}{(1+\lambda)^{k+1}}.$$

[提示: 利用 Stirling 公式 $n!=\sqrt{2\pi n}n^ne^{-n}(1+o(1))$].

7. 将 n 个带有号码 1 至 n 的球投入 n 个编有号码 1 至 n 的匣子, 并限制每一个匣子只能进一个球. 设球与匣子的号码一致的个数是 s_n 试证: $\frac{s_n-Es_n}{n}\to 0,$(P).

8. 试证 $\{\xi_n\}$ 服从大数定理的充要条件是

$$\lim_{n\to\infty}E\frac{\left(\sum\limits_{k=1}^{n}(\xi_k-E\xi_k)\right)^2}{n^2+\left(\sum\limits_{k=1}^{n}(\xi_k-E\xi_k)\right)^2}=0,$$

并由此推出马尔科夫大数定理.

9. 设 $\{\xi_n\}$ 为独立随机变数列， ξ_n 有分布密度为

$$f_n(x)=\frac{1}{\sqrt[4]{n}\cdot\sqrt{\pi}}\exp\left[-\frac{(x-\theta^n)^2}{\sqrt{n}}\right],$$

其中 $0<\theta<1$. 问 $\{\xi_n\}$ 是否服从强大数定理？

10. 设独立随机变数列 ξ_1: $\begin{pmatrix}1 & -1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$, ξ_n : $\begin{pmatrix}-\sqrt{n} & 0 & \sqrt{n}\\ \frac{1}{n} & 1-\frac{2}{n} & \frac{1}{n}\end{pmatrix}$, $n>1$. 试证 $\{\xi_n\}$ 服从强大数定律.

11. 设 X_n 独立同分布 $p\{X_n=2^{k-2\log k}\}=2^{-k}$ $(k=1,2,\cdots)$, 则大数定律成立.

12. 设 $\{X_n\}$ 是独立随机变数列， $DX_n=\frac{n+1}{\log(n+1)}$. 试问马尔科夫大数定律条件和柯尔莫哥洛夫定理条件是否满足？

(注意下列事实：若 $y_{n+1}>y_n$, $n=1,2,\cdots$, $\lim\limits_{n\to\infty}y_n=\infty$,

$$\lim_{n\to\infty}\frac{x_{n+1}-x_n}{y_{n+1}-y_n}\text{ 存在，则 }\lim_{n\to\infty}\frac{x_n}{y_n}=\lim_{n\to\infty}\frac{x_{n+1}-x_n}{y_{n+1}-y_n}\Bigg).$$

13. 设 $\{X_n\}$ 是相互独立且具有有限方差的随机变数列，若 $\sum\limits_{n=1}^{\infty}\frac{DX_n}{n^2}<\infty$, 则必有

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}DX_k=0.$$

14. 求证：如果对于独立随机变数列 ξ_n, 当 $A\to\infty$ 时有

$$\sup_{1\leqslant k<\infty}\int_{|x|\geqslant A}|x|\,dF_k(x)\to 0,$$

则 $\{\xi_n\}$ 服从大数定律.

15. 利用上题的结果证明，如果对独立随机变数列 ξ_n, 存在 $\alpha>1$ 及 $\beta>0$, 使 $E|\xi_n|^\alpha\leqslant\beta$, 则 $\{\xi_n\}$ 服从大数定律.

16. 设随机变数列 X_n 独立， X_n 在 $[-n,n]$ 上有均匀分布，问对 X_n 能否用中心极限定理.

17. 试问对下列独立随机变数列，李亚普诺夫定理是否成立，

(i) X_k : $\begin{pmatrix}-\sqrt{k} & \sqrt{k}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$;

(ii) X_k : $\begin{pmatrix}-k^\alpha & 0 & k^\alpha\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}$; $\alpha>0$.

18. 利用强大数定理证明：对任意 $q>p>0$,

$$\lim_{n\to\infty}\int_0^1\cdots\int_0^1\frac{x_1^q+\cdots+x_n^q}{x_1^p+\cdots+x_n^p}dx_1\cdots dx_n=\frac{p+1}{q+1}.$$

19. 求证：当 $n\to\infty$ 时， $e^{-n}\sum\limits_{k=0}^{n}\frac{n^k}{k!}\to\frac{1}{2}$.

20. 求证：当 $n\to\infty$ 时，

$$\frac{1}{\Gamma\left(\frac{n}{2}\right)}\sqrt{\left(\frac{n}{2}\right)^n}\int_0^{1+t\sqrt{\frac{2}{n}}}z^{\frac{n}{2}-1}e^{-\frac{nz}{2}}\,dz\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-\frac{z^2}{2}}\,dz.$$

21. 作 n 次广义贝努里试验，第 i 次成功概率是 p_i, 且 $\sum_{i=1}^{\infty} p_i q_i = \infty$. 设到 n 次试验为止，成功的次数是 s_n, 求证：

$$y_n = \frac{s_n - \sum_1^n p_i}{\sqrt{\sum_1^n p_i q_i}},$$

依分布趋于 $N(0,1)$.

22. 利用公式 $\binom{n}{k} p_n^k q_n^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$, $\lambda = \lim_{n\to\infty} np_n$, $q_n = 1 - p_n$. 计算

(i) 自一工厂的产品中任取 200 件，检查结果发现其中有四件废品，问我们能否相信此工厂出废品的概率 $\leqslant 0.005$.

(ii) 已知一工厂出废品的概率为 0.005, 任意取 1000 件，试求最小的 x, 使其中废品件数不超过 x 的概率 $\geqslant 0.999$.

(iii) 已知某种产品的废品率为 0.06, 问应取多少件产品，才能保证其中好产品的件数不少于 100 的概率 $\geqslant 0.92$.

23. 从装有 3 个白球与一个黑球的箱子中，还原地取 n 个球，设 m 是白球出现的次数，问 n 需要多大才能使得

$$P\left(\left|\frac{m}{n} - p\right| \leqslant 0.001\right) = 0.9964,$$

其中 p 是取每一次取到白球的概率.

24. 设 $\xi_1, \xi_2, \cdots$ 为独立随机变数列，而且一致有界，即存在常数 L, 使 $P(|\xi_k| \geqslant L) = 0$, 一切 k; 则当 $B_n^2 = \sum_{k=1}^{n} D\xi_k \to \infty$ $(n \to \infty)$ 时，中心极限定理成立.

提示：此时 ξ_k 的分布集中在 $[-L, L]$ 中. 取 $B_n > \frac{2L}{\tau}$, 则 §3.4(13) 中的积分域落于此区间之外，故每一积分都为 0 而麟德伯尔格条件成立.

25. 关于柯西分布：设 $\{\xi_k\}$ 独立，有相同的柯西分布密度 $f(x) = \frac{1}{\pi(1+x^2)}$, 因而特征函数为 $\varphi(t) = e^{-|t|}$. 由 §2.11(8) 与 (10), 知 $\frac{1}{n}\sum_{k=1}^{n} \xi_k$ 的特征函数也为 $e^{-|t|}$, 即与 ξ_k 同分布，故当 $n \to \infty$ 时，$\frac{1}{n}\sum_{k=1}^{n} \xi_k$ 不趋于 0, 因而相当于大数定理的结果不成立. 它也不服从中心极限定理，这不太奇怪，因为对柯西分布各级矩都不存在，但甚至有更强的结果：对任一列正实数 $\{d_n\}$, 可求出 $\frac{1}{d_n}\sum_{k=1}^{n} \xi_k$ 的特征函数为 $e^{-n|t|/d_n}$, 由 §2.5 中的常用分布表可见它对应的密度是 $\frac{d_n}{n\pi\left[1+\left(\frac{d_n x}{n}\right)^2\right]}$, 不趋于 $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

第四章 随机过程引论

§4.1 马尔科夫链

(一) **转移概率**. 设 $\{E_i\}(i=0,1,2,\cdots)$ 为一列随机试验，它们的一切可能出现的试验结果 ω_i 构成集 $\Omega=(\omega_0,\omega_1,\omega_2,\cdots)$, 于是每次试验后必出现 Ω 中某 ω_j. 本节中我们假定 Ω 是有穷集或可列集. 到现在为止，我们所考虑过的随机试验列几乎都是独立的：第 k 次试验的结果不依赖于前些次试验的结果. 精确地说，就是：对任意二正整数 m,k, 任意整数 $0\leqslant j_1<j_2<\cdots<j_l<m$, 有 [1]

$$P(\omega'_{m+k}\mid\omega'_m\omega'_{j_l}\cdots\omega'_{j_2}\omega'_{j_1})=P(\omega'_{m+k}) \tag{1}$$

其中 ω'_j 表事件："第 j 次试验出现 ω'_j"$(\omega'_j\in\Omega)$. 我们记得，贝努里试验就是由一列独立的随机试验组成的.

然而，实际中有许多随机试验列都不是独立的，因而有研究相依的 (即非独立的) 随机试验列的必要，其中最简单的一种由马尔科夫 (A. A. Марков) 首先研究.

称随机试验列 $\{E_i\}\,(i=0,1,2,\cdots)$ 为马尔科夫链 (简称马氏链), 如果对任意二正整数 m,k, 任意整数 $0\leqslant j_1<j_2<\cdots<j_l<m$, 有

$$P(\omega'_{m+k}\mid\omega'_m\omega'_{j_e}\cdots\omega'_{j_2}\omega'_{j_1})=P(\omega'_{m+k}\mid\omega'_m). \tag{2}$$

(2) 式的直观意义是：如已知现在 (第 m 次) 试验的结果 (ω'_m), 那么将来 (第 $m+k$ 次) 的结果 (ω'_{m+k}) 不依赖于过去第 j_k 次，$(k=1,\cdots,l)$ 的结果 $(\omega'_{j_l}\cdots\omega'_{j_2}\omega'_{j_1})$; 或者，简单些，如已知"现在", 则"将来"不依赖于"过去".

以 $p_{ij}(m,m+k)$ 表"在第 m 次试验出现 ω_i 的条件下，第 $m+k$ 次试验出现 ω_j"的条件概率，即 [2]

$$p_{ij}(m,m+k)=P(\omega'_{m+k}=\omega_j\mid\omega'_m=\omega_i), \tag{3}$$

称 $p_{ij}(m,m+k)$ 为转移概率. 一般说来，$p_{ij}(m,m+k)$ 不仅依赖于 i,j,k, 而且还依赖于 m; 如果它们与 m 无关，就称此马氏链为齐次的. 以后我们只讨论齐次马氏链，并将齐次二字省去. 特别重要的是一步转移概率

$$p_{ij}=p_{ij}(m,m+1), \tag{4}$$

它们构成一步转移概率矩阵 (简称转移概率矩阵)

$$P=(p_{ij})\;(i,j=0,1,2,\cdots). \tag{5}$$

[1] 这里及以后永远假定所用到的条件概率有意义，即构成条件的事件的概率大于 0.

[2] 下式中"$\omega'_m=\omega_i$"表事件"第 m 次试验出现的状态 ω'_m 为 ω_i", 也就是说，"第 m 次试验出现 ω_i", 以下同.

这矩阵具有下列二性质

$$0 \leqslant p_{ij} \leqslant 1, \tag{6}$$

$$\sum_j p_{ij} = 1. \tag{7}$$

实际上，由定义 $p_{ij} = P(\omega'_{m+1} = \omega_j \mid \omega'_m = \omega_i)$ 是条件概率，故 (6) 式是显然的；而 (7) 式则由于

$$\sum_j p_{ij} = \sum_j P(\omega'_{m+1} = \omega_j \mid \omega'_m = \omega_i) = P(\Omega \mid \omega'_m = \omega_i) = 1.$$

我们称任一具有性质 (6),(7) 的矩阵为随机矩阵. 类似地可见， n 步转移概率

$$p_{ij}(n) = p_{ij}(m, m+n) \tag{8}$$

的矩阵

$$P(n) = \big(p_{ij}(n)\big) \tag{9}$$

也是随机矩阵.

引理 1　对任二正整数 l, n, 有

$$p_{ij}(l+n) = \sum_k p_{ik}(l)\, p_{kj}(n); \tag{10}$$

亦即

$$P(l+n) = P(l)\, P(n) \tag{11}$$

证

$$\begin{aligned}
p_{ij}(l+n) &= P(\omega'_{m+l+n} = \omega_j \mid \omega'_m = \omega_i) \\
&= \frac{P(\omega'_m = \omega_i, \omega'_{m+l+n} = \omega_j)}{P(\omega'_m = \omega_i)} \\
&= \sum_k \frac{P(\omega'_m = \omega_i, \omega'_{m+l+n} = \omega_j, \omega'_{m+l} = \omega_k)}{P(\omega'_m = \omega_i, \omega'_{m+l} = \omega_k)} \cdot \frac{P(\omega'_m = \omega_i, \omega'_{m+l} = \omega_k)}{P(\omega'_m = \omega_i)} \\
&= \sum_k P(\omega'_{m+l+n} = \omega_j \mid \omega'_m = \omega_i, \omega'_{m+l} = \omega_k)\, P(\omega'_{m+l} = \omega_k \mid \omega'_m = \omega_i)
\end{aligned} \tag{12}$$

利用马氏性 (2) 与齐次性， $\sum\limits_k$ 号下第一因子等于

$$P(\omega'_{m+l+n} = \omega_j \mid \omega'_{m+l} = \omega_k) = p_{kj}(n),$$

第二因子等于 $p_{ik}(l)$, 故

$$p_{ij}(l+n) = \sum_k p_{ik}(l)\, p_{kj}(n). \square$$

公式 (10) 称为切普曼 - 柯尔莫各洛夫 (Chapman–Колмогоров) 方程. 利用此式，可以用一步的转移概率来表达多步的转移概率. 实际上

$$p_{ij}(2) = \sum_k p_{ik}\, p_{kj}.$$

一般地

$$p_{ij}(n+1)=\sum_k p_{ik}(n)\,p_{kj}=\sum_k p_{ik}\,p_{kj}(n). \tag{13}$$

由此可见，转移概率矩阵 (p_{ij}) 决定了随机试验列 $\{E_i\}$**转移** 过程的概率法则，就是说，如已知 $\omega'_m=\omega_i$, 那么 $\omega'_{m+n}=\omega_j$ 的概率就可自 (p_{ij}) 求出. 然而，(p_{ij}) 没有决定开始分布，亦即第 0 次 (即最初次) 试验中 $\omega'_0=\omega_i$ 的概率不能由 (p_{ij}) 表达，因而有必要引进

$$q_i=P(\omega'_0=\omega_i), \tag{14}$$

并称 $\{q_i\}=(q_0,q_1,q_2,\cdots)$ 为开始分布. 显然

$$q_i\geqslant 0,\ \sum_i q_i=1 \tag{15}$$

这种情况，正如微分方程决定运动的演变过程，而开始条件描绘运动的开始状态一样.

于是，一马氏链的概率法则完全由 $\{q_i\}$ 及 (p_{ij}) 所决定.

伴随着每一马氏链 $\{E_i\}$, 可以定义一列随机变数 $\xi_0,\xi_1,\xi_2,\cdots$, 如下：对每固定的 i, 令

$$\xi_i=j,\ 如\omega'_i=\omega_j; \tag{16}$$

就是说，如第 i 次试验出现 ω_j, 就定义 ξ_i 为 j, 即第 i 次试验的结果的足标. 因而每 ξ_i 只能取非负整数值，$\{\xi_i\}$ 记录了这一列试验的结果，称 $\{\xi_i\}$ 为马氏链 $\{E_i\}$ 的伴随随机变数列，通常也称 $\{\xi_i\}$ 为马氏链，并 $\{\xi_i\}$ 把的值域 $(0,1,2,\cdots)$ 称为此链的状态空间. 利用 $\{\xi_i\}$, 可把 (2) 式更详细地改写为

$$\begin{aligned}&P(\xi_{m+k}=i_{m+k}\mid\xi_m=i_m,\ \xi_{j_l}=i_{j_l},\cdots,\xi_{j_2}=i_{j_2},\ \xi_{j_1}=i_{j_1})\\&\quad=P(\xi_{m+k}=i_{m+k}\mid\xi_m=i_m),\end{aligned} \tag{17}$$

其中 $i_{m+k},i_m,i_{j_l},\cdots,i_{j_2},i_{j_1}$ 为任意非负整数，

(二) 例.

例 1 设 $\{E_i\},i=0,1,2,\cdots$ 为独立随机试验列，有相同的基本事件空间 $\Omega=(\omega_0,\omega_1,\omega_2,\cdots)$, 而且第 i 次试验中 E_i 出现 ω_j 的概率 p_j 与 i 无关，则 $\{E_i\}$ 是一马氏链. 实际上，由 (1) 可见 (2) 式两方的值都等于 $P(\omega'_{m+k})$ 而且

$$p_{ij}=P(\omega'_{m+1}=\omega_j\mid\omega'_m=\omega_i)=P(\omega'_{m+1}=\omega_j)=p_j$$

与 m 无关. 这链的转移概率矩阵为

$$P=(p_{ij})=\begin{pmatrix}p_0&p_1&p_2&\cdots\\p_0&p_1&p_2&\cdots\\p_0&p_1&p_2&\cdots\\\cdots&\cdots&\cdots&\cdots\end{pmatrix}.$$

特别，贝努里试验是一马氏链.

例 2 考虑 §1.5 例 6. 设质点 M 在整数点集 $(0,1,2,\cdots,a)$ 上作随机徘徊. 每经一单位时间按下列概率规则改变一次位置：如果它现在位于点 $j(0<j<a)$ 上，则下一步以概率

$p(0<p<1)$ 转移到 $j+1$, 以概率 q 到 $j-1, q=1-p$; 如果它现在在 0(或 a), 它以后就永远停留在 0(或 a). 我们把每次位移看成一随机试验 E_i, 把 j 看成 ω_j, 那么 $\{E_j\}$ 构成一马氏链，因为质点下一步转到某点 k 上的概率只依赖于它现在的位置而不依赖于过去的位置. 这时

$$P=\begin{pmatrix}1&0&0&0&\cdots&0&0&0\\q&0&p&0&\cdots&0&0&0\\0&q&0&p&\cdots&0&0&0\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&0&0&\cdots&q&0&p\\0&0&0&0&\cdots&0&0&1\end{pmatrix}$$

P 是 $a+1$ 级矩阵. 我们称 0 与 a 为吸引状态.

如果把 0 改为反射状态，就是说，质点如现在在 0, 下一步以概率 1 来到 1, 那么只须把 P 中第一横行改为 $(0,1,0,\cdots,0)$.

例 3 例 2 可如下一般化：设质点 M 在 $(\omega_0,\omega_1,\omega_2,\cdots)$ 上作随机运动，每经一单位时间改变一次位置，如果它现在在 ω_i, 那么它下一步转移到 ω_j 的概率为 p_{ij}, 这概率依赖于现在所在的位置 ω_i 而与过去无关，把每次位移看成一随机试验，于是 M 的一列位移构成一马氏链 $P=(p_{ij})$.

例 4 利用马氏链，可以求出某些偏微分方程的近似解. 考虑拉普拉斯方程

$$\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}=0 \tag{18}$$

及平面上某区域 G, 要寻求函数 $u(a),\big(a=(x,y)\big)$, 使 $u(a)$ 在 G 内满足 (18), 而在 G 的边界 Γ 上满足

$$u(a)=f(a),\big(a\in\Gamma\big) \tag{18$_1$}$$

这里 f 是已知函数.

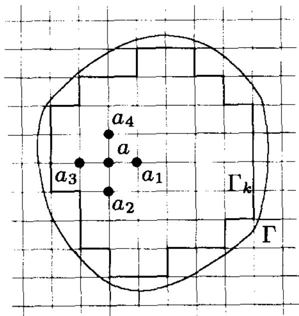


图 1

解微分方程的一种重要方法是所谓网络法，它使解微分方程变成解对应的差分方程，然而对差分方程的解取极限 (当网络边长 $\to 0$ 时) 即得微分方程的解.

在 G 上作长宽各为 h 的网络，它的交点称为结点，最接近 Γ 的结点构成集 Γ_h, G 中其余结点的集记为 G_h(图 1). 由微分方程的理论知对应于 (18) 的差分方程为

$$\left.\begin{aligned}&u(a)=\frac{1}{4}[u(a_1)+u(a_2)+u(a_3)+u(a_4)], &&\text{如}a\in G_h,\\&u(a)=f(a), &&\text{如}a\in\Gamma_h,\end{aligned}\right\} \tag{19}$$

其中 a_1,a_2,a_3,a_4 是 a 的四个相邻结点.

为解 (19), 考虑自 $a\in G_h$ 出发的随机运动的质点 M, 它下一步到达四邻点的概率各为 $\frac{1}{4}$, 再下一步又同样以 $\frac{1}{4}$ 的概率到达四邻点之一，如是继续，直至首次到达 Γ_h 时，便被吸引而停止运动. 以 $\xi(\in\Gamma_h)$ 表 M 首次到达时 Γ_h 所遇到的点，由于运动是随机的，ξ 是随

机点，于是得到随机变数 $f(\xi)$. 以 $v(a)$ 表 $f(\xi)$ 的数学期望，我们假定它存在. 可以证明，$v(a)$ 是差分方程 (19) 的解.

实际上，以 $p(a,b)$ 表自 a 出发的质点被吸引于 $b\in\Gamma_h$ 的概率. 如 $a\in G_h$ 则

$$p(a,b)=\frac{1}{4}\sum_{j=1}^{4}p(a_j,b)$$

$$v(a)=\sum_{b\in\Gamma_h}p(a,b)f(b)=\frac{1}{4}\sum_{j=1}^{4}\sum_{b\in\Gamma_h}p(a_j,b)f(b)=\frac{1}{4}\sum_{j=1}^{4}v(a_j).$$

故 $v(a)$ 在 G_h 内满足 (19) 中前一式. 如 $a\in\Gamma_h$ 则

$$p(a,b)=\begin{cases}1, & \text{如}a=b,\\ 0, & \text{如}a\neq b,\end{cases}$$

故

$$v(a)=\sum_{b\in\Gamma_h}p(a,b)f(b)=f(a),$$

因而 (19) 中后一式也满足.

注意 (19) 的解 $v(a)$ 其实依赖于 h, 当 h 充分小时，由微分方程知 $v(a)$ 近似于 (18) (18_1) 的解.

为了要造上述随机运动，可以利用计算机 (参看 §6.3) 或者简单地如下做；掷两枚硬币，以 (正、正)、(正、反)、(反、正)、(反、反) 分别表示东、南、西、北，每掷一次，就令 M 按掷出的方向移动一步. 自 a 出发，每到达一次 Γ_h, 就得一个数 $f=f(b)$; 重复 n 次后，得到 n 个数 $f_1,\cdots,f_n$. 根据强大数定理，当 n 充分大时，$\frac{1}{n}\sum\limits_{j=1}^{n}f_j$ 近似于 $v(a)$, 因而也近似于 (18),(18_1) 的解 $u(a)$. 这样，微分方程的近似解问题就化为作复合随机试验. 这种方法也是一种 Monte–Carlo 方法，关于这方面更详细的叙述见第六、七章.

(三) **遍历性.** 马氏链理论中一个重要的问题是：何时此链具有遍历性？称马氏链有遍历性，如果对一切 i,j, 存在不依赖于 i 的极限

$$\lim_{n\to\infty}p_{ij}(n)=p_j \tag{20}$$

这里 $p_{ij}(n)$ 是此链的 n 步转移概率.

(20) 式的直观意义是：不论质点自哪一个状态 ω_i 出发，当转移步数 n 充分大后，来到 ω_j 的概率都接近于 p_j. 因而反过来可以用 p_j 作为 $p_{ij}(n)$ 的近似值，只要 n 相当大.

这问题虽然已彻底解决，但要完全叙述却很冗长，下面的定理 1 给出了一个简单的充分条件以及求 p_j 的方法，我们只考虑有穷链 (即 Ω 只含有穷多个元，例如 k 个元的马氏链) 的情形.

定理 1 对有穷马氏链，如存在正整数 s, 使

$$p_{ij}(s)>0\ \ (\text{一切}\ i,j=1,\cdots,k), \tag{21}$$

则此链是遍历的；而且 (20) 中的 $\{p_1, p_2, \cdots, p_k\}$ 是方程组

$$p_j = \sum_{i=1}^{k} p_i p_{ij} \ (j = 1, \cdots, k) \tag{22}$$

的满足条件

$$p_j > 0, \ \sum_{j=1}^{k} p_j = 1 \tag{23}$$

的唯一解.

证 分成三步： (i) 先证 (20) 成立. 由 (10), 对 $n > 1$ 有

$$p_{ij}(n) = \sum_{l=1}^{k} p_{il} p_{lj}(n-1) \geqslant \min_{1 \leqslant l \leqslant k} p_{lj}(n-1) \sum_{l=1}^{k} p_{il} = \min_{1 \leqslant l \leqslant k} p_{lj}(n-1). \tag{24}$$

此式对一切 i 成立，故

$$\min_{1 \leqslant i \leqslant k} p_{ij}(n) \geqslant \min_{1 \leqslant i \leqslant k} p_{ij}(n-1), \tag{25}$$

从而存在极限

$$\lim_{n \to \infty} \min_{1 \leqslant i \leqslant k} p_{ij}(n) \geqslant \overline{p}_j \geqslant 0. \tag{26}$$

类似地证明

$$\max_{1 \leqslant i \leqslant k} p_{ij}(n) \leqslant \max_{1 \leqslant i \leqslant k} p_{ij}(n-1). \tag{27}$$

$$\lim_{n \to \infty} \max_{1 \leqslant i \leqslant k} p_{ij}(n) = \overline{\overline{p}}_j \geqslant 0. \tag{28}$$

如能证

$$\lim_{n \to \infty} \max_{1 \leqslant i, l \leqslant k} \left| p_{ij}(n) - p_{lj}(n) \right| = 0, \tag{29}$$

则 $\overline{p}_j = \overline{\overline{p}}_j$, 从而 (20) 成立，并且 $\overline{p}_i = \overline{p}_j = \overline{\overline{p}}_j$.

为证 (29), 取 $n > s$, 由 (10) 得

$$p_{ij}(n) = \sum_{r=1}^{k} p_{ir}(s) p_{rj}(n-s);$$

$$\begin{aligned} p_{ij}(n) - p_{lj}(n) &= \sum_{r=1}^{k} p_{ir}(s) p_{rj}(n-s) - \sum_{r=1}^{k} p_{lr}(s) p_{rj}(n-s) \\ &= \sum_{r=1}^{k} [p_{ir}(s) - p_{lr}(s)] p_{rj}(n-s). \end{aligned} \tag{30}$$

如 $p_{ir}(s) - p_{lr}(s) > 0$, 定义 $\alpha_{il}^{(r)} = p_{ir}(s) - p_{lr}(s)$; 如 $p_{ir}(s) - p_{lr}(s) \leqslant 0$, 定义 $\beta_{il}^{(r)} = p_{lr}(s) - p_{ir}(s)$.
由于

$$\sum_{r=1}^{k} p_{ir}(s) = \sum_{r=1}^{k} p_{lr}(s) = 1,$$

可见

$$\sum_{r=1}^{k}[p_{ir}(s)-p_{lr}(s)]=\sum_{(r)}\alpha_{il}^{(r)}-\sum_{(r)}\beta_{il}^{(r)}=0, \tag{31}$$

这里第一个 $\sum\limits_{(r)}$ 表对一切使 $\alpha_{il}^{(r)}$ 有定义的 r 求和，第二个 $\sum\limits_{(r)}$ 也有同样意义，于是

$$h_{il}=\sum_{(r)}\alpha_{il}^{(r)}=\sum_{(r)}\beta_{il}^{(r)}.$$

根据假定 (21), 对一切 l,r, 有 $p_{lr}(s)>0$, 故

$$\sum_{(r)}\alpha_{il}^{(r)}<\sum_{l=1}^{k}p_{il}(s)=1.$$

从而 $0\leqslant h_{il}<1$; 又因链是有穷的，知

$$0\leqslant h=\max_{1\leqslant i,l\leqslant k}h_{il}<1. \tag{32}$$

由 (30) 得

$$\begin{aligned}|p_{ij}(n)-p_{lj}(n)|&=\left|\sum_{(r)}\alpha_{il}^{(r)}p_{rj}(n-s)-\sum_{(r)}\beta_{il}^{(r)}p_{rj}(n-s)\right|\\&\leqslant\left|\max_{1\leqslant r\leqslant k}p_{rj}(n-s)\sum_{(r)}\alpha_{il}^{(r)}-\min_{1\leqslant r\leqslant k}p_{rj}(n-s)\sum_{(r)}\beta_{il}^{(r)}\right|\\&\leqslant h\left|\max_{1\leqslant r\leqslant k}p_{rj}(n-s)-\min_{1\leqslant r\leqslant k}p_{rj}(n-s)\right|\\&\leqslant h\max_{1\leqslant i,l\leqslant k}|p_{ij}(n-s)-p_{lj}(n-s)|,\end{aligned}$$

这式对一切 i,l 成立，故

$$\max_{1\leqslant i,l\leqslant k}|p_{ij}(n)-p_{lj}(n)|\leqslant h\max_{1\leqslant i,l\leqslant k}|p_{ij}(n-s)-p_{lj}(n-s)|.$$

利用此式 $\left[\frac{n}{s}\right]$ 次得

$$\begin{aligned}&\max_{1\leqslant i,l\leqslant k}|p_{ij}(n)-p_{lj}(n)|\\&\quad\leqslant h^{\left[\frac{n}{s}\right]}\max_{1\leqslant i,l\leqslant k}\left|p_{ij}\left(n-\left[\frac{n}{s}\right]s\right)-p_{lj}\left(n-\left[\frac{n}{s}\right]s\right)\right|\leqslant h^{\left[\frac{n}{s}\right]}.\end{aligned} \tag{33}$$

当 $n\to\infty$, 从而 $\left[\frac{n}{s}\right]\to\infty$ 时，由 (33) 得证 (29) 成立.

(ii) 次证 $\{p_1,p_2,\cdots,p_k\}$ 满足 (22) 与 (23). 由 (10)

$$p_{lj}(n+1)=\sum_{i=1}^{k}p_{li}(n)p_{ij} \tag{34}$$

对任意 l 成立，令 $n\to\infty$, 得知 $\{p_1,\cdots,p_k\}$ 满足 (22). 再在 $\sum\limits_{j=1}^{k} p_{lj}(n)=1$ 中，令 $n\to\infty$, 即得

$$\sum_{j=1}^{k} p_j = 1. \tag{35}$$

我们证明：对 **任一组** 满足 (22) 的 $p_1,\cdots,p_k$, 任意正整数 n, 有

$$p_j=\sum_{i=1}^{k} p_i p_{ij}(n),\ (j=1,\cdots,k) \tag{36}$$

实际上，当 $n=1$ 时化为 (22); 设 (36) 对 $n=m$ 成立，以 p_{jl} 乘 (36) 两边并对 j 求和，再利用 (22) 即得

$$\begin{aligned} p_l &= \sum_{j=1}^{k} p_j p_{jl} = \sum_{j=1}^{k}\sum_{i=1}^{k} p_i p_{ij}(m) p_{jl} \\ &= \sum_{i=1}^{k} p_i\, p_{il}(m+1), \quad (l=1,\cdots,k) \end{aligned}$$

故 (36) 对 $n=m+1$ 也成立.

显然，由 (20) 知 $p_j\geqslant 0$; 如果说有某 j 使 $p_j=0$, 在 (36) 中取 $n=s$, 得

$$0=\sum_{i=1}^{k} p_i p_{ij}(s).$$

但由 (21), 一切 $p_{ij}(s)>0$, 故为使上式成立，必须一切 $p_i=0\ (i=1,\cdots,k)$, 而这与 (35) 矛盾. 于是 (23) 成立.

(iii) 最后证 (20) 中极限是 (22)(23) 的唯一解. 设有某一组数 $\{v_1,\cdots,v_k\}$ 也满足 (22)(23). 由 (36) 得

$$v_j=\sum_{i=1}^{k} v_i p_{ij}(n)\,(i=1,\cdots,n).$$

令 $n\to\infty$, 并注意 (23) 对 $\{v_1,\cdots,,v_k\}$ 成立，得

$$v_j=\sum_{i=1}^{k} v_i p_j = p_j\sum_{i=1}^{k} v_i = p_j. \ \square$$

例 5 设马氏链只有三个状态，它的转移概率矩阵为

$$P=\begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix},$$

其中 $1>p>0, q=1-p$. 这矩阵给出如下的在 (0,1,2) 上的随机徘徊：自 0 出发，下一步停留在 0 的概率为 q, 来到 1 的概率为 p; 自 1 出发，到 0 及 2 的概率分别为 q 与 p; 自 2 出

发，停留在 2 及到 1 的概率各为 p 与 q. 对 $s=1$, 条件 (21) 虽不满足，但因

$$\left(p_{ij}^{(2)}\right)=P^2=\begin{pmatrix} q^2+pq & pq & p^2 \\ q^2 & 2pq & p^2 \\ q^2 & pq & pq+p^2 \end{pmatrix}$$

的元都大于 0, 故 (21) 对 $s=2$ 满足，从而此马氏链具有遍历性：

$$\lim_{n\to\infty} p_{ij}^{(n)}=p_j\ (i,j=1,2,3).$$

为求 $p_j, j=1,2,3$, 列出方程 (22):

$$\begin{aligned} p_1 &= p_1q+p_2q, \\ p_2 &= p_1p+p_3q, \\ p_3 &= p_2p+p_3p. \end{aligned}$$

由此解得 $p_2=\dfrac{p}{q}p_1, p_3=\left(\dfrac{p}{q}\right)^2 p_1$. 由 (23) $\sum\limits_{j=1}^{3} p_j=1$ 得

$$p_1\left[1+\frac{p}{q}+\left(\frac{p}{q}\right)^2\right]=1$$

如 $p=q=\frac{1}{2}$, 解得 $p_1=p_2=p_3=\frac{1}{3}$, 这表明在极限情形三状态是等可能的；如 $p\neq q$, 则得

$$p_j=\frac{1-\dfrac{p}{q}}{1-\left(\dfrac{p}{q}\right)^3}\left(\frac{p}{q}\right)^{j-1}\ (j=1,2,3).$$

例 6 作为遍历性不成立的平凡的例，考虑转移概率矩阵为 $P=\begin{pmatrix} 1 & 0 \\ 0, & 1 \end{pmatrix}$ 的马氏链. 由于 $P^n=P$, 可见

$$p_{11}(n)=p_{22}(n)=1, p_{12}(n)=p_{21}(n)=0.$$

由于 $\lim\limits_{n\to\infty} p_{11}(n)\neq \lim\limits_{n\to\infty} p_{21}(n)$, 故遍历性不成立.

§4.2 随机过程论中的基本概念

(一) 直观的例. 在实际许多问题中，我们不仅需要对随机现象作**一次**观察，而且要作多次，甚至要接连不断地观察它的演变过程，这种需要促进了随机过程论的诞生. 直观地说，随机过程论研究的对象正是随机现象演变过程的概率规律性. 先来看一些实际的例子：

例 1 医院不断地登记新生婴儿的性别，以 0 表“男”, 以 1 表“女”, 并以 ξ_n 表第 n 次登记的数字，ξ_n 是一变数，取值 1 或 0, 不断登记下去时，便得到一列，$\xi_1,\xi_2,\cdots$, 记为 $\{\xi_n, n=1,2,\cdots\}$.

例 2 以 ξ_t 表在时间 $(0,t]$ 内所见流星的个数，ξ_t 可取非负整数值，不断观察便得 $\{\xi_t, t\in[0,\infty)\}$.

例 3 1827 年布朗 (Brown) 发现水中花粉 (或其他液体中某微粒) 在不停地运动，后来称为布朗运动. 起因是由于花粉受到了水中分子的碰撞，每秒钟所受碰撞次数多到 10^{21} 次，这些随机的微小的碰撞力的总和使得花粉作随机运动. 以 ξ_t 表花粉在 t 时所在位置的一个坐标， ξ_t 可取一切实数值. 不断观察便得到 $\{\xi_t, t\in[0,\infty)\}$.

例 4 考虑纺织机所纺出的某一根棉纱，以 ξ_t 表 t 时纺出的横截面的直径，由于工作条件随 t 变化而起伏不同， ξ_t 一般也随 t 而变 (图 1), 于是得到 $\{\xi_t, t\in[0,\infty)\}$.

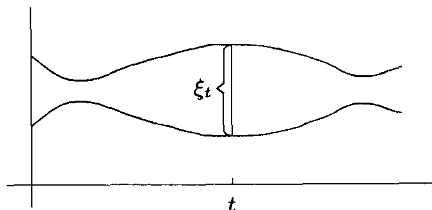


图 1

根据上面的实例，我们来下随机过程的严格数学定义：

设已给概率空间 $(\Omega,\mathcal{F},p)$ 及参数集 $T\subset(-\infty,\infty)$, 如对每一 $t\in T$, 有一定义在此空间上的随机变数 $\xi_t(\omega),(\omega\in\Omega)$, 就称 $\{\xi_t(\omega), t\in T\}$,(简记为 $\{\xi_t, t\in T\}$) 为随机过程 (有时简称为过程).

例如： n 维随机向量 $\{\xi_1,\xi_2,\cdots,\xi_n\}$ 是一随机过程，这里 $T=(1,2,\cdots,n)$; 随机变数列 $\{\xi_1,\xi_2,\cdots,\}$ 也是一随机过程， $T=(1,2,\cdots)$.

由此可见，随机过程是 n 维随机向量、随机变数列的一般化，它是随机变数 ξ_t 的集， $t\in T$. 回顾一下概率论研究对象的逐步扩大是有益的：从随机事件到随机变数到 n 维随机向量进而随机变数列 (它可看成为可列维随机向量), 以至随机过程.

正象每个随机变数有一分布函数一样，对每个随机过程 $\{\xi_t, t\in T\}$, 任取正整数 n 及 $t_i\in T, i=1,\cdots,n$, 令

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=P(\xi_{t_1}\leqslant x_1,\cdots,\xi_{t_n}\leqslant x_n).$$

则 $F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$ 是一 n 元分布函数，当 n 在正整数集中任意变动， $t_1,\cdots,t_n$ 在 T 中变动时，便得到许多 $F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$, 全体这种函数的集

$$\{F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)\}$$

称为 $\{\xi_t, t\in T\}$ 的有穷维分布函数族.

(二) 分类. 可以把随机过程如下分类. 对 $\{\xi_t, t\in T\}$, 以 Ξ 表 $\xi_t(t\in T)$ 的值域， $\Xi\subset R_1$. 随着 T,Ξ 是离散 (即有穷或可列) 集或非离散集而分成四类：

1° 参数集 T 离散、值域集 Ξ 离散的随机过程；

2° 参数集 T 离散、值域集 Ξ 非离散的随机过程；

3° 参数集 T 非离散、值域集 Ξ 离散的随机过程；

4° 参数集 T 非离散、值域集 Ξ 非离散的随机过程.

然而这种分类法是比较表面的，更深刻的是按过程的概率结构而分类，例如

i 如果对任一正整数 n, 任意 n 个不同的 $t_i\in T$, 随机变数 $\xi_{t_1},\cdots,\xi_{t_n}$ 是独立的，则称 $\{\xi_t, t\in T\}$ 为具独立随机变数的随机过程.

当 $T=(1,2,\cdots)$ 时，这类过程已在第三章中研究过，例 1 中的过程属于这种类型.

ii 如果对任一正整数 n, 任意 $t_i\in T, t_1<t_2<\cdots<t_n$, 随机变数

$$\xi_{t_2}-\xi_{t_1},\xi_{t_3}-\xi_{t_2},\cdots,\xi_{t_n}-\xi_{t_{n-1}}$$

是独立的，则称 $\{\xi_t, t\in T\}$ 为独立增量随机过程.

iii 马尔科夫随机过程 (定义见后).

iv 平稳随机过程 (定义见后).

等等.

一般是把上面两种分类法结合起来，例如，马尔科夫过程中分为 "T 离散、Ξ 离散"、"T 离散、Ξ 非离散" 等四种.

在过程论里流行着形象的语言，设想有一作随机运动的质点 M，以 ξ_t 表 M 在 t 时的位置，于是 $\{\xi_t, t\in T\}$ 描绘了 M 所作的随机运动的演变过程，因而通常称值域集 Ξ 为状态空间，Ξ 中的每一点 x 称为一 **状态**，把 T 看作时间 t 的集，并把 $\xi_t = x$ 形象地说成 "在 t 时运动质点 M 位于状态 x" 等等.

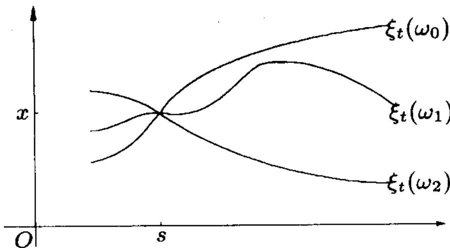


图 2

随机过程 $\xi_t(\omega)$ 可以看成为二元 $t\in T, \omega\in\Omega$ 的函数，当 $t\in T$ 固定时，$\xi_t(\omega)$ 是一随机变数；当 $\omega\in\Omega$ 固定时，作为 $t\in T$ 的函数 $\xi_t(\omega)$ 是一定义在 T 上的普通的函数，称为对应于基本事件 ω 的样本函数，或者，用运动的语言来说，它是可能的 (对应于 ω 的) 一种**轨道**. 当 ω 在 Ω 中变动时，便得到一切可能的轨道的集合. 图 2 中给出了 $\{\xi_t(\omega), t\in[0,\infty)\}$ 的三个轨道，它们在 s 时都经过点 x: $\xi_s = x$.

(三) **例**. 考虑两个重要的具体过程，它们分别是例 2 与例 3 中实际中的随机现象演变过程的抽象数学模型.

例 5 称随机过程 $\{\xi_t, t\in[0,\infty)\}$ 为普阿松过程，如果它是取非负整数值的独立增量过程，而且增量 $\xi_t-\xi_s (0\leqslant s<t)$ 有普阿松分布：

$$P(\xi_t-\xi_s=k)=e^{-\lambda(t-s)}\frac{[\lambda(t-s)]^k}{k!}, (k=0,1,2,\cdots) \tag{1}$$

其中 $\lambda>0$ 为常数.

如果假定例 2 中的流星流是一普阿松流，即满足 §2.4(三) 中条件 1° A!"2° A!"3° A!"4°，则由 §2.4 知例 2 中的 $\{\xi_t, t\in[0,\infty)\}$ 是普阿松过程.

它的样本函数是跃度为一的不减的阶梯函数 ($\xi_0(\omega)=0$ 时，见图 3)，τ_i 表第 i 个流星出现的时间.

例 6 称随机过程 $\{\xi_t, t\in[0,\infty)\}$ 为维纳 (Wiener) 过程，如果它是独立增量过程，而且增量 $\xi_t-\xi_s (0\leqslant s<t)$ 有正态分布 $N(0,\sigma\sqrt{t-s})$，其中 $\sigma>0$ 为常数.

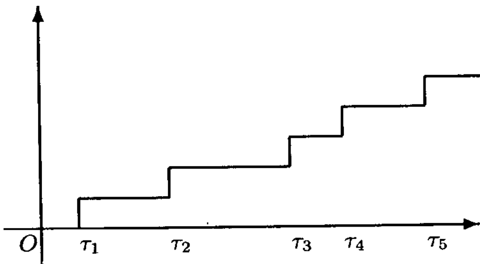


图 3

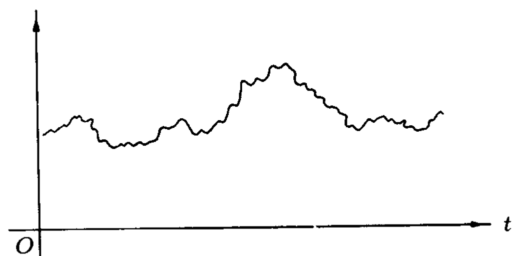


图 4

通常用维纳过程来模拟布朗运动，原因在于：设液体的质量是均匀的，由于微粒的运动是由许多分子碰撞所产生的许多小随机位移的和，自然可设想自时刻 s 到 t 的位移 $\xi_t-\xi_s$ 是许多几乎独立的小位移的和，故由中心极限定理，可假定 $\xi_t-\xi_s$ 有正态分布. 由液体的均匀性，还可设 $E(\xi_t-\xi_s)=0$，而且方差只依赖于时间区间长度 $t-s$，$D(\xi_t-\xi_s)=\sigma^2(t-s)$，其中 σ 是依赖于液体本身的扩散常数（不同的液体一般应有不同的 σ）. 最后，由于 $\xi_{t_2}-\xi_{t_1},\cdots,\xi_{t_n}-\xi_{t_{n-1}}$ 是对应于 **不相交** 区间的增量，它们分别为许多几乎独立的小位移的和，故可假定为独立的. 稍加条件后，可以证明维纳过程的样本函数以概率 1 是连续的，然而在任一固定 t 上以概率 1 不可微分. 图 4 给出了一个样本函数的示意图.

§4.3 马尔科夫过程

(一) 马尔科夫性. 一类重要的过程是马尔科夫过程，它是马氏链的一般化. 考虑随机过程 $\{\xi_t,t\in T\}$，状态空间记为 Ξ，设 $\Xi\in\mathcal{B}_1$，即设 Ξ 是一维波雷耳集. 对 T 中任意有穷多个点 s_i,s，及 $t,s_1\leqslant s_2\leqslant\cdots\leqslant s_n\leqslant s<t$，任意 Ξ 中的点 x_i,x，以

$$F(s_1,x_1,\cdots,s_n,x_n,s,x;t,A) \tag{1}$$

表在已知 $\xi_{s_1}=x_1,\cdots,\xi_{s_n}=x_n,\xi_s=x$ 的条件下，事件 $(\xi_t\in A)$ 的条件概率，其中 $A\in\mathcal{B}_1$，如果对于这一过程，(1) 中函数 F 不依赖于 $s_1,x_1,\cdots,s_n,x_n$，即如

$$F(s_1,x_1,\cdots,s_n,x_n,s,x;t,A)=F(s,x;t,A). \tag{2}$$

就称此过程为马尔科夫过程，简称马氏过程. (2) 式所表达的性质称为马尔科夫性（简称马氏性）.

也可解释 (2) 式为：如已知“现在：$\xi_s=x$”，则“将来：$\xi_t\in A$”不依赖于“过去：$\xi_{s_i}=x_i,i=1,\cdots,n$”；或者说，过程是无后效的（没有后果的）.

由 4.1(17) 可见，§4.1(16) 中定义的马氏链 $\{\xi_i\}$ 是 T 及 Ξ 都离散的马氏过程.

为方便起见，以下总设 $T=[0,\infty),\Xi=R_1$.

马氏过程的重要数字特征是 (2) 中的 $F(s,x;t,A)$，它是一四元函数，其中自变元为 $s\in T,x\in\Xi,t\in T,A\in\mathcal{B}_1,s<t$.

显然，仿照 §2.1 定理 2 的证明，可见

1° 当 s,x,t 固定时，$F(s,x;t,A)$ 作为集 A 的函数，是 $(R_1,\mathcal{B}_1)$ 上的概率测度.

因而，特别地，四元函数

$$F(s,x;t,y)=F\big(s,x;t,(-\infty,y]\big) \tag{3}$$

当 s,x,t 固定时，作为 $y\in R_1$ 的函数是一分布函数；即对 y 单调不减、右连续，而且

$$\lim_{y\to-\infty}F(s,x;t,y)=0,\quad \lim_{y\to\infty}F(s,x;t,y)=1. \tag{4}$$

以后我们还假定 $F(s,x;t,A)$ 具有下列二性质：

2° 当 s,t,A 固定时，$F(s,x;t,A)$ 是 x 的波雷耳可测函数；

3° $F(s,x;,t,A)$ 满足切普曼－柯尔莫各洛夫方程：对任意固定的 $0\leqslant s<\tau<t$ 及 x、A, 有

$$F(s,x;t,A)=\int_{-\infty}^{\infty} F(\tau,z;t,A)\,d_zF(s,x;\tau,z). \tag{5}$$

函数 $F(s,x;t,A)$ 只对 $s<t$ 有定义，当 $s=t$ 时，我们补定义

4°

$$F(s,x;s,A)=\delta(x,A)\equiv\begin{cases}0, & 如x\overline{\in}A,\\ 1, & 如x\in A.\end{cases}$$

当 $F(s,x;t,A)$ 具备性质 1°、2°、3°、4° 时，就称它为此马氏过程的转移概率，而由 (3) 定义的 $F(s,x;t,y)$ 称为过程转移分布函数.

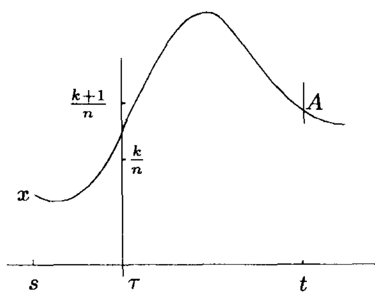


图 1

注 1 我们来解释一下这些假设的意义，以后会看到，(5) 式起着极其重要的作用. 在马氏链情况，(5) 式化为 §4.1(10)(积分号化为求和号 $\sum$), 故这时 3° 中的假设总是成立的. 为了保证 (5) 中积分有意义，应当假定被积函数 $F(\tau,z;t,A)$ 对 z 的可测性，这就是要引进假设 2° 的原因. 为了说明假设 3° 的直观意义，当 $F(\tau,z;t,A)$ 是 z 的连续函数时，可以这样想：质点在 s 时自 x 出发，要于 t 时转移到 A 中，可分成两步走 (见图 1):

(a) 在 τ 时通过某一区间 $\left(\frac{k}{n},\frac{k+1}{n}\right]$, 概率为 $F\left(s,x;\tau,\frac{k+1}{n}\right)-F\left(s,x;\tau,\frac{k}{n}\right)$;

(b) 在 τ 时自 $\left(\frac{k}{n},\frac{k+1}{n}\right]$ 中其一点 z_{nk} 出发，于 t 时转移到 A 中，概率为 $F(s,x,\tau,z_{nk};t,A)$. 当 n 充分大因而区间 $\left(\frac{k}{n},\frac{k+1}{n}\right]$ 的长度充分小时，可以指望

$$F(s,x;t,A)\approx\sum_{k=-\infty}^{\infty}F(s,x,\tau,z_{nk};t,A)\Big[F\Big(s,x;\tau,\frac{k+1}{n}\Big)-F\Big(s,x;\tau,\frac{k}{n}\Big)\Big],$$

利用马氏性 (2) 及上面假定的 $F(\tau,z;t,A)$ 对 z 的连续性，即得

$$\begin{aligned}F(s,x;t,A)&\approx\sum_{k=-\infty}^{\infty}F(\tau,z_{nk};t,A)\Big[F\Big(s,x;\tau,\frac{k+1}{n}\Big)-F\Big(s,x;\tau,\frac{k}{n}\Big)\Big]\\&\approx\int_{-\infty}^{\infty}F(\tau,z;t,A)d_zF(s,x;\tau,z).\end{aligned}$$

至于 4°, 它的概率意义是显然的：在同一时刻 s, 质点尚未离开 x.

如果马氏过程的转移概率 $F(s,x;t,A)$ 对 s,t 的依赖性只体现在对 $t-s$ 的依赖上，即如 $F(s,x;t,A)$ 只是 $t-s$,x 及 A 的函数时，就称此马氏过程是齐次的.

转移概率是马氏过程的重要特征，它在很大程度上决定此过程的运动进程，因此，在实际问题中，如果可以判断某过程是马氏过程，那么如何求出 $F(s,x;t,A)$ 就成为重要的问题了. 本节以下的内容就是围绕这一重要问题来叙述的.

对 $F(s,x;t,A)$ 进一步作种种不同的假定后，就得到种种不同的马氏过程，我们只讨论其中最重要的两种：跳跃型过程与扩散过程.

(二) **跳跃型马氏过程**. 设已给二函数 $q(t,x)$ 及 $\Pi(t,x,A)$, 自变元为 $t \geqslant 0, x \in R_1, A \in \mathcal{B}_1$, 满足下列条件:

(i) 对每固定的 $x, q(t,x)$ 是 t 的连续函数, 对每固定的 t, 是 x 的 $\mathcal{B}_1$ 可测函数; $q(t,x)$ 是 (t,x) 的非负有界函数.

(ii) 对每固定的 x 及 $A, \Pi(t,x,A)$ 对 t 连续; 对固定的 t,x, 它关于 A 是 $\mathcal{B}_1$ 上的概率测度; 对固定的 t,A, 是 x 的 $\mathcal{B}_1$ 可测函数, $\Pi(t,x,\{x\}) = 0$.

称马氏过程 $\{\xi_t, t \geqslant 0\}$ 为跳跃型的, 如存在如上的二函数 $q(t,x), \Pi(t,x,A)$, 使它的转移概率 $F(s,x;t,A)$, 当 $t-s>0$ 充分小时可表为

$$F(s,x;t,A) = \{1 - q(s,x)(t-s)\}\delta(x,A) + q(s,x)(t-s)\Pi(s,x,A) + o(t-s) \tag{6}$$

$$F(s,x;t,A) = \{1 - q(t,x)(t-s)\}\delta(x,A) + q(t,x)(t-s)\Pi(t,x,A) + o(t-s) \tag{7}$$

而且

$$\lim_{s \uparrow t} F(s,x;t,A) = \lim_{s \downarrow t} F(s,x;t,A) = \delta(x,A) \tag{8}$$

(6)(7) 式表达了"跳跃型"的直观意义: (6) 式表示, 当 $t-s$ 充分小时, 在 (s,t) 中 ξ_t 以概率 $\{1-q(s,x)(t-s)\} + o(t-s)$ 保持不变而等于 x, 另一可能是以概率 $q(s,x)(t-s) \times \Pi(s,x,A) + o(t-s)$ 转移到 A 中. (7) 式也同样解释.

例 1 设 $\{\xi_t, t \geqslant 0\}$ 为普阿松过程, $\xi_0 \equiv 0$. 由增量的独立性, 如已知 $\xi_s = i$, 则 $\xi_t - \xi_s = \xi_t - i (t > s)$ 与一切 $\xi_u = \xi_u - \xi_0$ 独立 $(0 \leqslant u \leqslant s)$, 因而在 $\xi_s = i$ 的条件下, ξ_t 与 ξ_u 独立, $(t > s, 0 \leqslant u \leqslant s)$. 因此知此过程是马氏过程: 转移概率为

$$F(s,i;t,\{j\}) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^{j-i}}{(j-i))!}, (j \geqslant i) \tag{9}$$

注意 $\sum\limits_{j \geqslant i} F(s,i;t,\{j\}) = 1$, 故 $F(s,i;t,\{k\}) = 0, (k < i)$. 又因 $F(s,i;t,\{j\})$ 只依赖于 $t-s$, 可见它还是齐次马氏过程. 由 (9) 得

$$F(s,i;t,\{j\}) = \{1 - \lambda \cdot (t-s)\}\delta(i,\{j\}) + \lambda(t-s)\delta(i+1,\{j\}) + o(t-s). \tag{10}$$

故过程是跳跃型的, $q(s,i) \equiv \lambda$, 又 $\Pi(s,i,\{j\}) = 1$ 或 0, 视 $i+1=j$ 或 $i+1 \neq j$ 而定.

以 $\Pi(t,x,y)$ 表 $\Pi(t,x,A)$ 的分布函数, 即

$$\Pi(t,x,y) = \Pi(t,x,(-\infty,y]).$$

定理 1 跳跃型过程的转移概率 $F(s,x;t,A)$ 满足下面两个微积分方程

$$\frac{\partial F(s,x;t,A)}{\partial s} = q(s,x)\left[F(s,x;t,A) - \int_{-\infty}^{\infty} F(s,y;t,A) d_y \Pi(s,x,y)\right] \tag{11}$$

$$\frac{\partial F(s,x;t,A)}{\partial t} = -\int_A q(t,y) d_y F(s,x;t,y) + \int_{-\infty}^{\infty} q(t,y)\Pi(t,y,A) d_y F(s,x;t,y). \tag{12}$$

方程 (11) 中左方是对后一参数 s 求导数, 故称为向后方程 (当时 $s=0$, 理解 $\frac{\partial F(s,x;t,A)}{\partial s}$ 为右导数); 而 (12) 称为向前方程. 这二方程都由柯尔莫哥洛夫得到.

证 对 $\Delta s>0, s<s+\Delta s<t$ 由 (5) 得

$$F(s,x;t,A)=\int_{-\infty}^{\infty}F(s+\Delta s,y;t,A)d_yF(s,x;s+\Delta s,y). \tag{13}$$

按 (6) 展开 $F(s,x;s+\Delta s,y)$, 代入上式得

$$\begin{aligned}F(s,x;t,A)=&\int_{-\infty}^{\infty}F(s+\Delta s,y;t,A)d_y\big[1-q(s,x)\Delta s+o(\Delta s)\big]\delta(x,y)\\&+\int_{-\infty}^{\infty}F(s+\Delta s,y;t,A)d_y\big[q(s,x)\Delta s+o(\Delta s)\big]\Pi(s,x,y),\end{aligned}$$

这里 $\delta(x,y)=0$ 或 1, 视 $y<x$ 或 $y\geqslant x$ 而定，它是 $\delta(x,A)$ 的分布函数. 由于分布 $\delta(x,A)$ 集中在一点 x 上，故

$$\int_{-\infty}^{\infty}F(s+\Delta s,y;t,A)d_y\delta(x,y)=F(s+\Delta s,x;t,A),$$

从而

$$\begin{aligned}F(s,x;t,A)=&\big[1-q(s,x)\Delta s\big]F(s+\Delta s,x;t,A)\\&+\Delta sq(s,x)\int_{-\infty}^{\infty}F(s+\Delta s,y;t,A)d_y\Pi(s,x,y)+o(\Delta s),\end{aligned}$$

于是

$$\begin{aligned}\frac{F(s+\Delta s,x;t,A)-F(s,x;t,A)}{\Delta s}=&\,q(s,x)F(s+\Delta s,x;t,A)\\&-q(s,x)\int_{-\infty}^{\infty}F(s+\Delta s,y;t,A)d_y\Pi(s,x,y)+o(1),\end{aligned}$$

令 $\Delta s\to 0$, 并注意 $\Delta s>0$, 就得

$$\frac{\partial^+F(s,x;t,A)}{\partial s}=q(s,x)\left[F(s,x;t,A)-\int_{-\infty}^{\infty}F(s,y;t,A)d_y\Pi(s,x,y)\right], \tag{14}$$

这里 $\frac{\partial^+}{\partial s}$ 表右导数. 如果我们用

$$F(s-\Delta s,x;t,A)=\int F(s,y;t,A)d_yF(s-\Delta s,x;s,y)$$

代替 (13), 以 (7) 代替 (6), 类似地可证 (14) 的 $\frac{\partial^+F(s,x;t,A)}{\partial s}$ 换为左导数 $\frac{\partial^-F(s,x;t,A)}{\partial s}$ 后也成立，这便证明了 (11).

为了证明 (12), 仍由 (5), 对 $\Delta t>0$, 有

$$F(s,x;t+\Delta t,A)=\int_{-\infty}^{\infty}F(t,y;t+\Delta t,A)d_yF(s,x;t,y). \tag{15}$$

由于

$$F(s,x;t,A)=\int_{-\infty}^{\infty}\delta(y,A)d_yF(s,x;t,y),$$

可改写 (15) 为

$$\frac{1}{\Delta t}[F(s,x;t+\Delta t,A)-F(s,x;t,A)]$$
$$=\int_{-\infty}^{\infty}\frac{F(t,y;t+\Delta t,A)-\delta(y,A)]}{\Delta t}d_yF(s,x;t,y).$$

令 $\Delta t\to 0$, 并利用 (6) 及积分有界收敛定理 (这里本质上用到 $q(t,y)$ 的有界性), 得知 (12) 成立, 但其中 $\frac{\partial}{\partial t}$ 应理解为右导数 $\frac{\partial^+}{\partial t}$; 类似地可证对 $\frac{\partial^-}{\partial t}$ 也成立, 故 (12) 得证. □

例 2 设 $\{\xi_t,t\geqslant 0\}$ 为跳跃型马氏过程, 但只有可列 (或有穷) 多个状态 $0,1,2,\cdots$. 这时习惯上用 $i,j,k,\cdots$ 等来表示状态, 并把 $F(s,i;t,\{k\})$ 及 $q(s,i)$ 分别记成 $p_{ik}(s,t)$ 与 $q_i(s)$, 把 $q(s,i)\times\Pi(s,i,\{k\})$ 记为 $q_{ik}(s)$. 注意由 (ii), $q_{ii}(s)=0$. 方程 (11)(12) 化为两个偏微分方程组

$$\frac{\partial p_{ik}(s,t)}{\partial s}=q_i(s)p_{ik}(s,t)-\sum_{j\neq i}q_{ij}(s)p_{jk}(s,t)\,(i,k=0,1,2,\cdots); \tag{16}$$

$$\frac{\partial p_{ik}(s,t)}{\partial t}=-p_{ik}(s,t)q_k(t)+\sum_{j\neq k}p_{ij}(s,t)q_{jk}(t),\ (i,k=0,1,2,\cdots). \tag{17}$$

特别, 如过程是齐次的, 即如 $p_{ik}(s,t)=p_{ik}(t-s)$ 只依赖于差 $t-s$ 时, 应设 $q_i(s)=q_i, q_{ij}(s)=q_{ij}$ 不依赖于时间 s. 令 $u=t-s$, 注意

$$\frac{\partial p_{ik}(s,t)}{\partial s}=\frac{\partial p_{ik}(t-s)}{\partial s}=-\frac{\partial p_{ik}(t-s)}{\partial(t-s)}=-p'_{ik}(u),$$

可见 (16)(17) 分别化为常微分方程组 [1]

$$p'_{ik}(u)=-q_ip_{ik}(u)+\sum_{j\neq i}q_{ij}p_{jk}(u),\ (i,k=0,1,2,\cdots); \tag{18}$$

$$p'_{ik}(u)=-p_{ik}(u)q_k+\sum_{j\neq k}p_{ij}(u)q_{jk},\ (i,k=0,1,2,\cdots). \tag{19}$$

开始条件为:

$$p_{ij}(0)=\begin{cases}1, & i=j,\\ 0, & i\neq j.\end{cases}$$

对于例 1 中普阿松过程, (18)、(19) 化为

$$p'_{ik}(u)=-\lambda p_{ik}(u)+\lambda p_{i+1k}(u), \tag{20}$$

---

[1] 如令 $q_{ii}=-q_i$, 则 (18)(19) 及开始条件分别化为

$$P'(u)=QP(u), P'(u)=P(u)Q, P(0)=I\,(\text{单位矩阵}),$$

其中 $Q=(q_{ij}), P(u)=\big(p_{ij}(u)\big), P'(u)=\big(p'_{ij}(u)\big)$ 为矩阵. 由此知状态个数有穷时解为 $P(u)=e^{Qu}=\sum\limits_{n=0}^{\infty}\frac{Q^nu^n}{n!}$. 又对普阿松过程, (20) 或 (21) 在此开始条件下之解为

$$p_{ik}(u)=e^{-\lambda u}\frac{(\lambda u)^{k-i}}{(k-i)!},\ (k\geqslant i);=0,\ (k\leqslant i).$$

$$p'_{ik}(u) = -\lambda p_{ik}(u) + \lambda p_{ik-1}(u). \tag{21}$$

在许多具体问题中，实践的经验会告诉我们某随机现象的演变过程可看成一跳跃型的马氏过程，而且 (6) 中的 $q(s,x)$ 及 $\Pi(s,x,A)$ 可以由观察得到．这时为了寻求转移概率 $F(s,x;t,A)$, 由定理 1, 只要解方程 (11)(或 (12))．当然，我们应该寻求这样的解 $F(s,x;t,A)$, 它对固定的 s,x,t, 关于 A 是 $\mathcal{B}_1$ 上的概率测度．于是发生两个重要的理论问题：

a. 这样的解是否存在？

b. 如存在，又是否唯一？

费勒 (W. Feller) 证明了：在条件 (i) (ii) 下，这样的解是唯一存在的 (如果除去 (i) 中的 $q(t,x)$ 有界性条件，那么解不是唯一的); 并且找到了此解的表达式．于是这解就是所要求的转移概率．

(三) **扩散过程**．另一类重要的马氏过程是扩散过程．这类过程起源于物理学中对微粒的随机扩散运动 (例如布朗运动) 的研究．它不同于跳跃型过程，跳跃型过程有下列特点：在任一很短的时间内，质点以很大的概率 **不发生** 位移，但如一旦发生，则位移 **很大**(出现跳跃), 所以可以想象运动的轨道以概率 1 是不连续的，轨道是阶梯函数．扩散过程则相反．在任一很短时间内，质点都可能 **发生** 位移，然而位移 **很小**，因而可以想象在一定条件下扩散过程的轨道以概率 1 是连续函数．

设 $\{\xi_t, t \geqslant 0\}$ 为马氏过程，以 $F(s,x;t,y)$ 表它的转移分布函数．称此过程为扩散过程，如果 $F(s,x;t,y)$ 满足下列三条件 (a),(b),(c):

对任意 $\delta > 0, \Delta t > 0$, 有

(a)

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| \geqslant \delta} d_y F(t,x;t+\Delta t,y) \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| \geqslant \delta} d_y F(t-\Delta t,x;t,y) = 0;\end{aligned} \tag{22}$$

(b)

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| < \delta} (y-x) d_y F(t,x;t+\Delta t,y) \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| < \delta} (y-x) d_y F(t-\Delta t,x;t,y) = a(t,x);\end{aligned} \tag{23}$$

(c)

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| < \delta} (y-x)^2 d_y F(t,x;t+\Delta t,y) \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| < \delta} (y-x)^2 d_y F(t-\Delta t,x;t,y) = b(t,x).\end{aligned} \tag{24}$$

我们先对这些条件作些解释：

1. 条件 (a) 前式表示：于 t 时自 x 出发的质点，经 Δt 后跑出 x 的邻域 $(x-\delta, x+\delta)$ 的概率是比 Δt 更高级的无穷小，不论区间的半径 $\delta > 0$ 如何：这反映质点在很短时间内不能得到很大的位移．

2. 在条件 (a) 下，条件 (b) 中第一、二项中极限不依赖于 $\delta(\delta>0)$, 所以 $a(t,x)$ 与 δ 无关；(c) 中极限也如此. 实际上，简记 $\frac{1}{\Delta t}\int_{|y-x|<\delta}(y-x)d_yF(t,x;t+\Delta t,y)$ 为 $\frac{1}{\Delta t}\int_{|y-x|<\delta}(y-x)$, 则如 $0<\delta_1<\delta_2$, 由 (a) 有

$$\begin{aligned}\left|\frac{1}{\Delta t}\int_{|y-x|<\delta_1}(y-x)-\frac{1}{\Delta t}\int_{|y-x|<\delta_2}(y-x)\right|&=\left|\frac{1}{\Delta t}\int_{\delta_1\leqslant|y-x|<\delta_2}(y-x)\right|\\ \leqslant\frac{\delta_2}{\Delta t}\int_{|y-x|\geqslant\delta_1}d_yF(t,x;t+\Delta t,y)&\to 0,\ (\Delta t\to 0).\end{aligned}\tag{25}$$

对 (23)(24) 中其它三极限也可类似证明.

3. 为了说明 (b),(c) 中 $a(t,x),b(t,x)$ 的概率意义，我们把条件 (a) 加强为

(a′) 对任意 $\delta>0,\Delta t>0$, 有

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|\geqslant\delta}(y-x)^2d_yF(t,x;t+\Delta t,y)\\ &=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|\geqslant\delta}(y-x)^2d_yF(t-\Delta t,x;t,y)=0,\end{aligned}\tag{26}$$

则由 (23)(24) 得 [1]

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{-\infty}^{\infty}(y-x)d_yF(t,x;t+\Delta t,y)=a(t,x),\tag{27}$$

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{-\infty}^{\infty}(y-x)^2d_yF(t,x;t+\Delta t,y)=b(t,x).\tag{28}$$

注意当 $\xi_t=x$ 时,

$$\int_{-\infty}^{\infty}(y-x)d_yF(t,x;t+\Delta t,y)=E[\xi_{t+\Delta t}-\xi_t],$$

$$\int_{-\infty}^{\infty}(y-x)^2d_yF(t,x;t+\Delta t,y)=E[\xi_{t+\Delta t}-\xi_t]^2.$$

故 (27)(28) 式表示 $a(t,x)$ 为在 t 时自 x 出发的质点的瞬时平均速度，而 $b(t,x)$ 则与瞬时平均动能成比例.

例 3 设 $\{\xi_t,t\geqslant 0\}$ 为维纳过程， $\xi_0\equiv 0$, 象例 1 中证明一样，可知它是马氏过程，又

$$F(s,x;t,y)=\frac{1}{\sigma\sqrt{2\pi(t-s)}}\int_{-\infty}^{y}e^{-\frac{(z-x)^2}{2\sigma^2(t-s)}}dz.\tag{29}$$

条件 (a)(b)(c) 都满足，这时

$$a(t,x)\equiv 0,\quad b(t,x)\equiv\sigma^2,\tag{30}$$

所以它是扩散过程.

现在来推导向后与向前方程.

定理 2 设 $F(s,x;t,y)$ 为扩散过程的转移分布函数，如果偏导数

$$\frac{\partial F(s,x;t,y)}{\partial x},\quad\frac{\partial^2F(s,x;t,y)}{\partial x^2}$$

---

[1] 根据布尼亚科夫斯基不等式，由 (26) 知 (26) 中的 $(y-x)^2$ 换为 $(y-x)$ 后仍成立.

存在而且对 s,x,y 及 $t(>s)$ 连续，则 $F(s,x;t,y)$ 满足向后方程

$$\frac{\partial F(s,x;t,y)}{\partial s}=-a(s,x)\frac{\partial F(s,x;t,y)}{\partial x}-\frac{b(s,x)}{2}\frac{\partial^2 F(s,x;t,y)}{\partial x^2}. \tag{31}$$

证 由 (15) 得对 $\Delta s>0$, 有

$$\begin{aligned}&\frac{F(s,x;t,y)-F(s-\Delta s,x;t,y)}{-\Delta s}\\&=\frac{1}{\Delta s}\int_{-\infty}^{\infty}\big[F(s,z;t,y)-F(s,x;t,y)\big]d_zF(s-\Delta s,x;s,z)\\&=\frac{1}{\Delta s}\int_{|z-x|\geqslant\delta}\big[F(s,z;t,y)-F(s,x;t,y)\big]d_zF(s-\Delta s,x;s,z)\\&\quad+\frac{1}{\Delta s}\int_{|z-x|<\delta}\big[F(s,z;t,y)-F(s,x;t,y)\big]d_zF(s-\Delta s,x;s,z).\end{aligned} \tag{32}$$

由 (a), 右方第一项当 $\Delta s\to 0$ 时趋于 0; 而第二项根据假设及泰勒 (Taylor) 展式化为

$$\begin{aligned}&\frac{\partial F(s,x;t,y)}{\partial x}\frac{1}{\Delta s}\int_{|z-x|<\delta}(z-x)d_zF(s-\Delta s,x;s,z))\\&+\frac{1}{2}\cdot\frac{\partial^2 F(s,x;t,y)}{\partial x^2}\frac{1}{\Delta s}\int_{|z-x|<\delta}\{(z-x)^2+o\big[(z-x)^2\big]\}d_zF(s-\Delta s,x;s,z),\end{aligned}$$

当 $\Delta s\to 0$ 时，由 (b)(c) 上式趋于

$$a(s,x)\frac{\partial F(s,x;t,y)}{\partial x}+\frac{b(s,x)}{2}\frac{\partial^2 F(s,x;t,y))}{\partial x^2}.$$

因而得证 (31) 中以 $\frac{\partial^-}{\partial s}$ 换 $\frac{\partial}{\partial s}$ 后成立；利用定理 1 证明中理由，类似知对 $\frac{\partial^+}{\partial s}$ 也成立，从而 (31) 得证. □

以下设转移密度

$$f(s,x;t,y)=\frac{\partial F(s,x;t,y)}{\partial y}$$

存在，由 (31) 得

$$\frac{\partial f(s,x;t,y)}{\partial s}=-a(s,x)\frac{\partial f(s,x;t,y)}{\partial x}-\frac{b(s,x)}{2}\frac{\partial^2 f(s,x;t,y)}{\partial x^2}. \tag{33}$$

定理 3 设扩散过程的转移分布密度 $f(s,x;t,y)$ 存在，下列偏导数

$$\frac{\partial f(s,x;t,y)}{\partial t},\quad \frac{\partial}{\partial y}\big[a(t,y)f(s,x;t,y\big],\quad \frac{\partial^2}{\partial y^2}\big[b(t,y)f(s,x;t,y)\big]$$

也存在而且连续，又设 (23)(24) 中收敛对 x 是均匀的，则 $f(s,x;t,y)$ 满足向前方程

$$\frac{\partial f(s,x;t,y)}{\partial t}=-\frac{\partial}{\partial y}\big[a(t,y)f(s,x;t,y)\big]+\frac{1}{2}\frac{\partial^2}{\partial y^2}\big[b(t,y)f(s,x;t,y)\big] \tag{34}$$

证 任取二数 $a,b(a<b)$, 及非负连续函数 $R(y)$, 它有二阶连续导数，并且

$$R(y)=0,\ \text{如}y<a\ \text{或}y>b.$$

由连续性得

$$R(a)=R(b)=R'(a)=R'(b)=R''(a)=R''(b)=0. \tag{35}$$

注意

$$\int_a^b \frac{\partial f(s,x;t,y)}{\partial t}R(y)dy=\lim_{\Delta t\to 0}\int_{-\infty}^{\infty}\frac{f(s,x;t+\Delta t,y)-f(s,x;t,y)}{\Delta t}R(y)dy. \tag{36}$$

由 (5) 有

$$f(s,x;t+\Delta t,y)=\int_{-\infty}^{\infty}f(s,x;t,z)f(t,z;t+\Delta t,y)dz,$$

代入 (36), 改变积分次序，然后把 y,z 对调:

$$\begin{aligned}
&\int_a^b \frac{\partial f(s,x;t,y)}{\partial t}R(y)dy\\
&=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(s,x;t,z)f(t,z;t+\Delta t,y)R(y)dzdy-\int_{-\infty}^{\infty}f(s,x;t,y)R(y)dy\right]\\
&=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left[\int_{-\infty}^{\infty}f(s,x;t,z)\int_{-\infty}^{\infty}f(t,z;t+\Delta t,y)R(y)dydz-\int_{-\infty}^{\infty}f(s,x;t,y)R(y)dy\right]\\
&=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{-\infty}^{\infty}f(s,x;t,y)\left[\int_{-\infty}^{\infty}f(t,y;t+\Delta t,z)R(z)dz-R(y)\right]dy.
\end{aligned}$$

根据泰勒展式

$$R(z)=R(y)+(z-y)R'(y)+\frac{(z-y)^2}{2}R''(y)+o\left[(z-y)^2\right],$$

由于 $R(z)$ 有界及 (a)

$$\int_{|y-z|\geqslant\delta}f(t,y;t+\Delta t,z)R(z)dz=o(\Delta t),$$

$$\int_{|y-z|<\delta}f(t,y;t+\Delta t,z)dz=1+o(\Delta t),$$

故

$$\begin{aligned}
&\int_{-\infty}^{\infty}f(t,y;t+\Delta t,z)R(z)dz-R(y)=\int_{-\infty}^{\infty}\left[R(z)-R(y)\right]f(t,y;t+\Delta t,z)dz\\
&=R'(y)\int_{|y-z|<\delta}(z-y)f(t,y;t+\Delta t,z)dz\\
&\quad+\frac{1}{2}R''(y)\int_{|y-z|<\delta}\left[(z-y)^2+o(z-y)^2\right]f(t,y;t+\Delta t,z)dz+o(\Delta t).
\end{aligned}$$

于是

$$\begin{aligned}
&\int_a^b \frac{\partial f(s,x;t,y)}{\partial t}R(y)dy\\
&=\lim_{\Delta t\to 0}\int_{-\infty}^{\infty}f(s,x;t,y)\Big\{R'(y)\int_{|y-z|<\delta}(z-y)f(t,y;t+\Delta t,z)dz\\
&\quad+\frac{1}{2}R''(y)\int_{|y-z|<\delta}\left[(z-y)^2+o(z-y)^2\right]f(t,y;t+\Delta t,z)dz+o(\Delta t)\Big\}dy.
\end{aligned}$$

令 $\Delta t \to 0$, 根据 (23)(24) 中收敛的均匀性，上式右方化为

$$\int_{-\infty}^{\infty} f(s,x;t,y)\big[a(t,y)R'(y)+\frac{1}{2}b(t,y)R''(y)\big]dy.$$

又因 $R'(y)=R''(y)=0$ ($y \leqslant a$, 或$y \geqslant b$), 故

$$\int_a^b \frac{\partial f(s,x;t,y)}{\partial t}R(y)dy=\int_a^b f(s,x;t,y)\big[a(t,y)R'(y)+\frac{1}{2}b(t,y)R''(y)\big]dy. \tag{37}$$

由分部积分

$$\int_a^b f(s,x;t,y)a(t,y)R'(y)dy=-\int_a^b R(y)\frac{\partial}{\partial y}\big[a(t,y)f(s,x;t,y)\big]dy,$$

$$\int_a^b f(s,x;t,y)b(t,y)R''(y)dy=\int_a^b R(y)\frac{\partial^2}{\partial y^2}\big[b(t,y)f(s,x;t,y)\big]dy,$$

代入 (37), 移项后即得

$$\int_a^b \Big\{\frac{\partial f(s,x;t,y)}{\partial t}+\frac{\partial}{\partial y}\big[a(t,y)f(s,x;t,y)\big]-\frac{1}{2}\frac{\partial^2}{\partial y^2}\big[b(t,y)f(s,x;t,y)\big]\Big\}R(y)dy=0. \tag{38}$$

由此及 $R(y)$ 的任意性即可推出 (34). 实际上，设 (34) 不成立，因而 (38) 式积分号下大括号中表达式在某定点 (s_0,x_0,t_0,y_0) 上不为 0. 由假定这式中诸项连续，故存在 α,β, 使 $y_0\in(\alpha,\beta)$ 而且此式在 (α,β) 中不变号，由 a,b 的任意性可设 $a<\alpha,\beta<b$. 今取一如上的 $R(y)$, 并使它在 (α,β) 中大于 0 而在 (α,β) 外等于 0. 对这样选定的 $R(y)$,(38) 左方积分不为 0, 于是发生矛盾，故 (34) 成立. □

像 (二) 中一样，在实际问题中并不是要验证 $F(s,x;t,y)$ 或 $f(s,x;t,y)$ 是否满足向后或向前方程，而是根据比较容易找到的 $a(t,x)$ 及 $b(t,x)$ 来求 $F(s,x;t,y)$ 或 $f(s,x;t,y)$, 于是需要研究向后及向前方程解的存在性与唯一性问题. 在对 $a(t,x),b(t,x)$ 加一些条件后，费勒曾证明满足 (a),(b),(c) 的转移密度唯一存在.

例 4 设 $a(t,x)\equiv 0, b(t,x)=\sigma^2>0$,(31),(34) 分别化为

$$\frac{\partial F(s,x;t,y)}{\partial s}=\frac{-\sigma^2}{2}\frac{\partial^2 F(s,x;t,y)}{\partial x^2},$$

$$\frac{\partial f(s,x;t,y)}{\partial t}=\frac{\sigma^2}{2}\frac{\partial^2 f(s,x;t,y)}{\partial y^2}.$$

不难直接验证，下列函数是它们的解

$$F(s,x;t,y)=\frac{1}{\sigma\sqrt{2\pi(t-s)}}\int_{-\infty}^{y}\exp\Big\{-\frac{(u-x)^2}{2\sigma^2(t-s)}\Big\}du,$$

$$f(s,x;t,y)=\frac{1}{\sigma\sqrt{2\pi(t-s)}}\exp\Big\{-\frac{(y-x)^2}{2\sigma^2(t-s)}\Big\},$$

于是得到维纳过程的转移分布函数与转移密度.

*§4.4 独立增量过程

(一) 定义. 称 $\{\xi_t, t \geqslant 0\}$ 为独立增量过程，如果对任意正整数 n, 任意 $0 \leqslant t_1 < t_2 < \cdots < t_n$ 增量

$$\xi_{t_2} - \xi_{t_1}, \xi_{t_3} - \xi_{t_2}, \cdots, \xi_{t_n} - \xi_{t_{n-1}} \tag{1}$$

是相互独立的随机变数. 称独立增量过程为齐次的，如果对任意 $s \geqslant 0, \tau > 0$, 增量

$$\xi_{s+\tau} - \xi_s$$

的分布函数只依赖于 τ 而不依赖 s.

普阿松过程与维纳过程都是齐次独立增量过程.

例 1 设 $\{\eta_n\}, n = 0, 1, 2, \cdots$ 是一列独立随机变数，定义

$$\xi_n = \sum_{k=0}^{n} \eta_k. \tag{2}$$

则 $\{\xi_n\}$ 是一以 $T = \{0, 1, 2, \cdots\}$ 为参数集的独立增量过程，当而且只当 $\{\eta_n\}$ 有相同分布时，$\{\xi_n\}$ 是齐次的.

反之，设 $\{\xi_n\}, n = 0, 1, 2, \cdots$ 是独立增量过程，$\xi_0 = 0$, 定义

$$\eta_n = \xi_n - \xi_{n-1}, (n > 0). \tag{3}$$

显然，$\{\eta_n\}$ 是独立随机变数列；当而且只当 $\{\xi_n\}$ 为齐次时，$\{\eta_n\}$ 有相同的分布.

由此可见：对离散的参数集 T, 任一满足 $\xi_0 = 0$ 的 (齐次) 独立增量过程可表为独立 (同分布) 随机变数列的部分和.

在独立增量过程的研究中，一个重要的问题是求出增量 $\xi_{s+\tau} - \xi_s$ 的分布函数的普遍形式. 我们已经知道，对于普阿松过程，$\xi_{s+\tau} - \xi_s$ 有普阿松分布 $P(\lambda\tau)$; 而对维纳过程，则有正态分布 $N(0, \sigma\sqrt{\tau})$.

这问题已经圆满地解决，为了叙述简单起见，我们以后永远假定 $\{\xi_t, t \geqslant 0\}$ 满足下列条件：

(a) 它是齐次的；

(b) $m_2(t) = E(\xi_{s+t} - \xi_s)^2 < \infty$; 一切 $t \geqslant 0$;

(c) $m_1(t) = E(\xi_{s+t} - \xi_s) = 0$, 一切 $t \geqslant 0$.

并称满足条件 (a)、(b)、(c) 的独立增量过程为标准独立增量过程.

我们指出，条件 (c) 不是本质的，因为如 (c) 不满足，我们可以考虑过程 $\{\xi_t - E\xi_t, t \in [0, \infty)\}$.

(二) 基本定理. 以 $F(x, \tau)$ 及 $f(t, \tau)$ 分别表 $\xi_{s+\tau} - \xi_s$ 的分布函数与特征函数. 我们来证明下列基本定理.

定理 1 设 $\{\xi_t, t \geqslant 0\}$ 为标准独立增量过程，则 [1]

$$\log f(t, \tau) = -\frac{1}{2}\sigma_0^2 \tau t^2 + \tau \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} dG(x), \tag{4}$$

[1](4) 中被积函数 $\frac{e^{itx}-1-itx}{x^2}$ 在 $x = 0$ 按连续性补定义为 $-\frac{t^2}{2}$, $\log f$ 取主值.

其中 $\sigma_0^2 \geqslant 0$ 为常数， $G(x)$ 是有界、右连续单调不减函数，而且在点 $x=0$ 连续；其次，如有另一对如此的 $\widetilde{\sigma}_0^2$ 及 $\widetilde{G}(x)$ 满足 (4), 则

$$\sigma_0^2 = \widetilde{\sigma}_0^2, \quad G(x) - \widetilde{G}(x) \equiv b \text{ (常数)} .$$

在证明以前，先来看一些特殊情形.

1° 设 $\sigma_0^2 > 0, G(x) \equiv a$(常数); 则 (4) 化为

$$\log f(t,\tau) = -\frac{1}{2}\sigma_0^2 \tau t^2. \tag{5}$$

这表示 $\xi_{s+\tau} - \xi_s$ 有正态分布 $N(0, \sigma_0\sqrt{\tau})$, 从而 $\{\xi_t, t \geqslant 0\}$ 是维纳过程.

2° 设 $\sigma_0 = 0, G(x) = \lambda c^2 \varepsilon(x-c)$, 其中 $\varepsilon(x-c) = 0$ 或 1, 视 $x < c$ 或 $x \geqslant c$ 而定，又 $\lambda > 0, c \neq 0$ 为二常数. 这时 (4) 化为

$$\log f(t,\tau) = \lambda\tau(e^{ict} - 1 - ict).$$

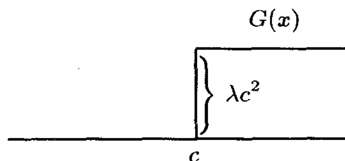


图 1 $\lambda c^2\varepsilon(x-c)$ 的图

这表明 $\xi_{s+\tau} - \xi_s + \lambda c\tau$ 的特征函数为

$$\exp\left[\lambda\tau(e^{ict} - 1)\right]. \tag{6}$$

它是下列离散分布

$$\begin{pmatrix} 0 & c & 2c & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}; \qquad p_k = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!} \quad (7)$$

的特征函数 (证仿 §2.11 例 2). 称此分布为广普阿松分布，并记为 $P(\lambda, c)$; 当 $c=1$ 时化为普阿松分布.

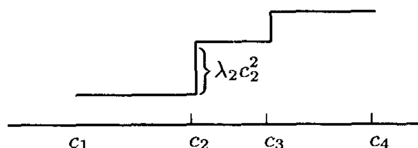


图 2 $\sum\limits_{i=1}^{n} \lambda_i c_i^2 \varepsilon(x-c_i)$ 的图

3° 设 $\sigma_0^2 > 0$, $G(x) = \sum\limits_{k=1}^{n} \lambda_k c_k^2 \varepsilon(x - c_k)$, 其中 $c_1 < c_2 < \cdots < c_n, c_k \neq 0, \lambda_k > 0$ 都为常数. 这时 (4) 化为

$$\log\, f(t,\tau) = -\frac{1}{2}\sigma_0^2\tau t^2 + \tau \cdot \sum_{k=1}^{n} \lambda_k(e^{ic_k t} - 1 - ic_k t) \tag{8}$$

这说明 $\xi_{s+\tau} - \xi_s + \tau\sum\limits_{k=1}^{n}\lambda_k c_k$ 的分布是正态分布 $N(0, \sigma_0\sqrt{\tau})$ 与 n 个广普阿松分布 $P(\lambda_k, c_k)(k = 1, \cdots, n)$ 的卷积.

定理 1的证明 分四步:

(A) 由增量的独立性及

$$\xi_{s+\tau_1+\tau_2} - \xi_s = (\xi_{s+\tau_1} - \xi_s) + (\xi_{s+\tau_1+\tau_2} - \xi_{s+\tau_1}), \tag{9}$$

知 $\xi_{s+\tau_1+\tau_2} - \xi_s$ 是二独立随机变数的和，故前者的分布函数是后二者的分布函数的卷积. 由齐次性得

$$F(x, \tau_1 + \tau_2) = F(x, \tau_1) * F(x, \tau_2); \tag{10}$$

$$f(t,\tau_1+\tau_2)=f(t,\tau_1)f(t,\tau_2). \tag{11}$$

由 (11) 得

$$f(t,1)=\left[f\left(t,\frac{1}{n}\right)\right]^n,\quad f\left(t,\frac{1}{n}\right)=\left[f(t,1)\right]^{\frac{1}{n}},$$

$$f\left(t,\frac{m}{n}\right)=f\left(t,\frac{1}{n}\right)^m=\left[f(t,1)\right]^{\frac{m}{n}}.$$

由此及特征函数的连续性，像 §2.4(9) 式的证明一样，可知对任意 $\tau>0$, 有

$$f(t,\tau)=\left[f(t,1)\right]^{\tau}. \tag{12}$$

试证 $f(t,\tau)\neq 0$, 因而 $\log f(t,\tau)$ 有意义. 实际上，再由 (11) 得 $f(t,\tau)=\left[f\left(t,\frac{\tau}{n}\right)\right]^n$, 故 $\left[f(t,\tau)\right]^{\frac{1}{n}}$ 是一特征函数. 令

$$g(t,\tau)=\lim_{n\to\infty}\left[f(t,\tau)\right]^{\frac{1}{n}}.$$

由于 $f(0,\tau)=1$ 及特征函数 $f(t,\tau)$ 的连续性，存在 $a>0$, 使 $f(t,\tau)\neq 0, |t|<a$, 于是 $g(t,\tau)=1, |t|<a$. 这说明特征函数列 $\left\{[f(t,\tau)]^{\frac{1}{n}}\right\}$ 的极限 $g(t,\tau)$ 在 $|t|<a$ 连续，由 §3.2 知 $g(t,\tau)$ 也是特征函数，从而它对 $t\in R_1$ 连续. 但由定义 $g(t,\tau)$ 只能取值 0 或 1, 故 $g(t,\tau)\equiv 1$(一切$t\in R_1$); $f(t,\tau)\neq 0$, 一切$t\in R_1, \tau>0$.

根据等式

$$\log a=\lim_{\delta\to\infty}\frac{a^\delta-1}{\delta}$$

及 (12) 得

$$\begin{aligned}\log f(t,\tau)&=\tau\log f(t,1)=\tau\lim_{\Delta\tau\to 0}\frac{f(t,1)^{\Delta\tau}-1}{\Delta\tau}\\&=\tau\lim_{\Delta\tau\to 0}\frac{f(t,\Delta\tau)-1}{\Delta\tau}.\end{aligned} \tag{13}$$

下面的证明路线是想以 $F(x,\Delta\tau)$ 来代替 $f(t,\Delta\tau)$, 然后令 $\Delta\tau\to 0$ 而取极限以得 (4). 为此先作些准备.

(B) 由假定 (b) 及 §2.11(一)(V), 可对 t 微分 (12) 两次而得

$$f''(t,\tau)=\tau(\tau-1)f^{\tau-2}(t,1)f'^2(t,1)+\tau f^{\tau-1}(t,1)f''(t,1).$$

取 $t=0$ 并注意由假定 (c)

$$f'(0,\tau)=iE(\xi_{s+\tau}-\xi_s)=0. \tag{14}$$

可见 $\xi_{s+\tau}-\xi_s$ 的方差 $D(\tau)$ 为

$$D(\tau)=\tau D(1)=\tau\sigma^2, \tag{15}$$

其中常数 $\sigma^2=D(1)\geqslant 0$. 如果 $\sigma^2=0$, 则 (4) 式对 $\sigma_0=0, G(x)\equiv b$(b 为任意常数) 显然成立，故以下设 $\sigma^2>0$.

(C) 由定义

$$f(t,\Delta\tau)=Ee^{it(\xi_{s+\Delta\tau}-\xi_s)}=\int_{-\infty}^{\infty}e^{itx}dF(x,\Delta\tau).$$

由 (14), $\int_{-\infty}^{\infty} xdF(x,\Delta\tau)=0$ 故

$$\frac{1}{\Delta\tau}\left[f(t,\Delta\tau)-1\right]=\frac{1}{\Delta\tau}\int_{-\infty}^{\infty}(e^{itx}-1-itx)dF(x,\Delta\tau). \tag{16}$$

由假定 (b) 可引进辅助函数

$$H(x,\Delta\tau)=\frac{1}{\Delta\tau}\int_{-\infty}^{x}\xi^2 dF(\xi,\Delta\tau). \tag{17}$$

显然 $H(x,\Delta\tau)$ 对 x 单调不减，右连续而且由 (15)

$$\lim_{x\to-\infty}H(x,\Delta\tau)=0,\quad \lim_{x\to\infty}H(x,\Delta\tau)=\sigma^2. \tag{18}$$

由 (13)(16)(17) 得

$$\log f(t,\tau)=\tau\lim_{\Delta\tau\to 0}\int_{-\infty}^{\infty}\frac{(e^{itx}-1-itx)}{x^2}dH(x,\Delta\tau).$$

取一列 $\Delta_n\tau\to 0$, 运用海来第一定理于 $H(x,\Delta_n\tau)$ 后，可见存在单调不减右连续函数 $H(x)$ 使某子列 $H(x,\Delta_{n_v}\tau)$ 在 $H(x)$ 的连续点上收敛于 $H(x)(v\to\infty)$, 再由海来第二定理 (§3.7), 知

$$\log f(t,\tau)=\tau\int_{-\infty}^{\infty}\frac{(e^{itx}-1-itx)}{x^2}dH(x). \tag{19}$$

显然

$$H(-\infty)\equiv\lim_{x\to-\infty}H(x)\geqslant 0;\quad H(\infty)\equiv\lim_{x\to\infty}H(x)\leqslant\sigma^2.$$

我们证明，上二式中等号成立. 实际上，由 (19), 当 t 充分小时

$$\log f(t,\tau)=-\frac{1}{2}\tau t^2\{H(\infty)-H(-\infty)\}+o(t^2);$$

另一方面，把 $f(t,\tau)$ 在 $\tau=0$ 按泰勒展开，并利用 (14)(15), 得

$$f(t,\tau)=1-\frac{1}{2}\theta\sigma^2t^2\tau,\ (|\theta|\leqslant 1)$$

于是

$$\log f(t,\tau)=-\frac{1}{2}\sigma^2t^2\tau+o(t^2).$$

与上面的结果比较，即得

$$H(-\infty)=0, H(\infty)=\sigma^2. \tag{20}$$

今定义 σ_0^2 及 $G(x)$:

$$0\leqslant\sigma_0^2=H(0)-H(0-)\leqslant\sigma^2, \tag{21}$$

$$G(x)=H(x)-\sigma_0^2\cdot\varepsilon(x).$$

则 $G(x)$ 有界右连续、单调不减，而且在 $x=0$ 连续

$$G(-\infty)=0.\ G(\infty)=\sigma_1^2=\sigma^2-\sigma_0^2. \tag{22}$$

利用此 σ_0^2 及 $G(x)$, 可改写 (19) 式为 (4).

(D) 下证 (4) 中 $G(x)$ 及 σ_0^2 的唯一性. 由假定 (b), 可把 (19) 对 t 微分两次而得

$$\int_{-\infty}^{\infty} e^{itx}dH(x)=-\frac{1}{\tau}\frac{\partial^2}{\partial t^2}\log f(t,\tau).$$

但 $\frac{H(x)}{\sigma^2}$ 是一分布函数，它应由自己的特征函数决定，故 $H(x)$, 因而 $G(x)$, 除差一常数被加项外，被 $f(t,\tau)$ 唯一决定 [从而附带证明了上述子列 $\{H(x,\Delta_{n_v}\tau)\}$ 的极限 $H(x)$(它还满足 (20)) 不依赖于 $\{n_v\}$ 的选择，即 $\lim\limits_{\Delta\tau\to 0}H(x,\Delta\tau)=H(x)$, 只要 x 是 H 的连续点]. 再由 (21) 即得 σ_0^2 的唯一性. □

我们对定理 1 中的 $G(x)$ 作些解释：由 (17)(19), 对 $x<0$, 当 $\Delta\tau\to 0$ 时，由海来第二定理，有

$$\frac{F(x,\Delta\tau)}{\Delta\tau}=\int_{-\infty}^{x}\frac{dH(\xi,\Delta\tau)}{\xi^2}\to\int_{-\infty}^{x}\frac{dG(\xi)}{\xi^2}=\varphi_1(x), \tag{23}$$

对 $x>0$

$$\frac{1-F(x,\tau)}{\Delta\tau}=\int_{x}^{\infty}\frac{dH(\xi,\Delta\tau)}{\xi^2}\to\int_{x}^{\infty}\frac{dG(\xi)}{\xi^2}=\varphi_2(x), \tag{24}$$

只要 x 是 $G(x)$ 的连续点. 改写上二式为

$$F(x,\Delta\tau)=\varphi_1(x)\Delta\tau+o(\Delta\tau)\ (x<0);$$

$$1-F(x,\Delta\tau)=\varphi_2(x)\Delta\tau+o(\Delta\tau)\ (x>0).$$

亦即在充分小的长为 $\Delta\tau$ 的区间中，对增量 $\xi_{s+\Delta\tau}-\xi_s$, 有

$$P(\xi_{s+\Delta\tau}-\xi_s\leqslant x)=\varphi_1(x)\Delta\tau+o(\Delta\tau)\quad(x<0),$$
$$P(\xi_{s+\Delta\tau}-\xi_s> x)=\varphi_2(x)\Delta\tau+o(\Delta\tau)\quad(x>0).$$

这二式说明了 $\varphi_1(x)$ 及 $\varphi_2(x)$ 的概率意义：增量不超过 $x(x<0)$ 的概率约为 $\varphi_1(x)\Delta\tau$, 大于 $x(x>0)$ 的概率约为 $\varphi_2(x)\Delta\tau$, 误差为一比 $\Delta\tau$ 高级的小量.

从例 1 自然想到，标准独立增量过程也许可看成独立同分布的随机变数的部分和，不过参数集 T 要从离散集 $\{0,1,2,\cdots\}$ 过渡到全体非负实数集，因而“求和”要理解为“积分”. 这样便自然猜想：标准独立增量过程也应具有独立，同分布、方差有穷的随机变数的和的性质. 作为中心极限定理的类似结果，我们来证明下面的定理 2. 为简单起见，不妨设

$$\xi_0\equiv 0 \tag{25}$$

(否则考虑过程 $\{\xi_t-\xi_0,\ t\geqslant 0\}$). 像 §4.3 中证明普阿松过程是马氏过程一样，可见满足 (25) 的独立增量过程是马氏过程. 由于 $\xi_\tau=\xi_\tau-\xi_0$, 知 ξ_τ 的分布函数与特征函数分别为 $F(x,\tau),f(t,\tau)$, 方差为 $\tau\sigma^2$, 期望为 0, 与 §3.4(2) 中标准化部分和相当的是随机变数

$$\frac{\xi_\tau}{\sqrt{D(\tau)}}=\frac{\xi_\tau}{\sigma\sqrt{\tau}}.$$

以下总设 $\sigma > 0$.

定理 2 标准独立增量过程 $\{\xi_t, t \geqslant 0\}$ 如满足条件 (25), 又 $E\xi_t^2 = D(t) > 0\,(t > 0)$, 则关于 x 均匀地有

$$\lim_{\tau\to\infty} P\left(\frac{\xi_\tau}{\sigma\sqrt{\tau}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}dy. \tag{26}$$

证 记 $\frac{\xi_\tau}{\sigma\sqrt{\tau}}$ 的特征函数为 $\overline{f}(t,\tau)$, 则

$$\overline{f}(t,\tau) = f\left(\frac{t}{\sigma\sqrt{\tau}}, \tau\right).$$

以 $\frac{t}{\sigma\sqrt{\tau}}$ 代入 (4) 中的 t, 得

$$\begin{aligned}\log \overline{f}(t,\tau) &= -\frac{\sigma_0^2}{2\sigma^2}t^2 + \tau\int_{-\infty}^{\infty}\frac{1}{x^2}\left[e^{\frac{itx}{\sigma\sqrt{\tau}}} - 1 - \frac{itx}{\sigma\sqrt{\tau}}\right]dG(x)\\ &= -\frac{t^2}{2} + \frac{\sigma^2-\sigma_0^2}{2\sigma^2}t^2 + \tau\int_{-\infty}^{\infty}\frac{1}{x^2}\left[e^{\frac{itx}{\sigma\sqrt{\tau}}} - 1 - \frac{itx}{\sigma\sqrt{\tau}}\right]dG(x).\end{aligned}$$

由 (22), $\sigma^2 - \sigma_0^2 = \int_{-\infty}^{\infty} dG(x)$, 代入上式得

$$\log \overline{f}(t,\tau) = -\frac{t^2}{2} + \tau\int_{-\infty}^{\infty}\frac{1}{x^2}\left[e^{\frac{itx}{\sigma\sqrt{\tau}}} - 1 - \frac{itx}{\sigma\sqrt{\tau}} - \frac{1}{2}\left(\frac{itx}{\sigma\sqrt{\tau}}\right)^2\right]dG(x).$$

因为被积号下方括号中的表达式当 $\tau\to\infty$ 时与 $\tau^{-\frac{3}{2}}$ 同级，故

$$\lim_{\tau\to\infty}\log\overline{f}(t,\tau) = -\frac{t^2}{2};$$

$$\lim_{\tau\to\infty}\overline{f}(t,\tau) = e^{-\frac{t^2}{2}}.\square$$

§4.5 平 稳 过 程

(一) **平稳性**. 另一类重要的过程是平稳过程，这类过程在无线电技术和自动控制等方面有着广泛的应用. 先来考察 §4.2 中例 4, 随着时间的演变，工作条件不断发生变化 (原料的质量、机器的性能、人的工作态度、自然条件如温度等的改变等等), 纺出纱的横截面直径自然会有波动. 但如**工作条件基本上稳定**, 没有剧烈的变化，则当我们同时观察 n 根纱 $\omega_1, \omega_2, \cdots, \omega_n$, 并以 $\xi_t(\omega_i)$ 表纱 ω_i 在 t 时的横截面的直径时，如果用 $G_t(x)$ 来记在 t 时横截面直径不超过定数 x 的纱的根数与 n 的比，即

$$G_t(x) = \frac{\text{满足 “}\xi_t(\omega_i)\leqslant x\text{” 的}i\text{ 的个数}}{n}. \tag{1}$$

那么，对充分大的 n, 就会发现 [1]: 对任意实数 τ,

$$G_t(x) \approx G_{t+\tau}(x).$$

[1] 下面的记号 “≈” 表 “近似”

就是说，这个比值在不同的时间基本上是不变的 (或者说，稳定的). 更一般地，有

$$G_{t_1,\cdots,t_n}(x_1,\cdots,x_n) \approx G_{t_1+\tau,\cdots,t_n+\tau}(x_1,\cdots,x_n), \tag{2}$$

其中

$$G_{s_1,\cdots,s_n}(x_1,\cdots,x_n) = \frac{\text{满足 “}\xi_{s_1}(\omega_i) \leqslant x_1,\cdots,\xi_{s_n}(\omega_i) \leqslant x_n\text{” 的}i\text{ 的个数}}{n}. \tag{3}$$

这种思想引导到一般的定义：

设 $\{\xi_t(\omega), t \in R_1\}$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机过程. 如果对任意正整数 n, 任意实数 $t_i, x_i (i = 1, 2, \cdots, n)$ 及 τ, 有

$$P(\omega : \xi_{t_1}(\omega) \leqslant x_1, \cdots, \xi_{t_n}(\omega) \leqslant x_n) = P(\omega : \xi_{t_1+\tau}(\omega) \leqslant x_1, \cdots, \xi_{t_n+\tau}(\omega) \leqslant x_n), \tag{4}$$

就称它为平稳过程.

全体函数

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) = P(\xi_{t_1} \leqslant x_1,\cdots,\xi_{t_n} \leqslant x_n) \tag{5}$$

构成此过程的有穷维分布函数族 (§4.2), 所以上面的定义也可以转述为：如果随机过程的一切有穷维分布函数不随时间的推移而改变，那么就称此过程是平稳过程.

对任一随机过程，可以象对随机变数一样来定义它的各阶矩 (如果存在的话).

$$E\xi_{t_1}^{m_1} \cdots \xi_{t_n}^{m_n} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{m_1} \cdots x_n^{m_n} d_{x_1,\cdots,x_n} F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) \tag{6}$$

特别

$$E\xi_{t+\tau}\xi_t = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 d_{x_1,x_2} F_{t,t+\tau}(x_1,x_2). \tag{7}$$

对于平稳过程 $\{\xi_t, t \in R_1\}$, 由于

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) = F_{t_1+\tau,\cdots,t_n+\tau}(x_1,\cdots,x_n), \tag{8}$$

所以如果假定 $E\xi_{t+\tau}\xi_t$ 存在，那么由 (7)(8) 可见

$$B(\tau) \equiv E\xi_{t+\tau}\xi_t = E\xi_\tau\xi_0 \tag{9}$$

与 t 无关，而只依赖于 τ. 换句话说，二阶矩不随时间推移而变，其实这对各阶矩都正确，只要它存在.

二阶矩 $E\xi_{t+\tau}\xi_t (t \in R_1, \tau \in R_1)$ 只能粗浅地刻划过程的性质，远远不能决定有穷维分布函数族；正如我们所知的，随机变数的各阶矩一般都不能决定分布函数一样. 虽然如此，在许多实际问题中，往往只需要过程的二阶矩性质 (即能通过二阶矩来表达的性质), 也往往因为二阶矩远比有穷维分布函数族容易观察到，所以二阶矩有时能起到重要的作用，这样就产生了弱平稳过程的概念.

对随机过程 $\{\xi_t, t \in R_1\}$, 如果 (9) 式成立，即如 $B(\tau) = E\xi_{t+\tau}\xi_t$ 不依赖于 t, 就称这过程为弱平稳过程.

我们指出，平稳过程未必是弱平稳的，因为它的 $B(\tau)$ 不一定存在；弱平稳过程更未必是平稳的，因为由 (9) 一般不能推出 (8).

弱平稳的概念可以推广到复随机过程．称定义在概率空间 $(\Omega,\mathcal{F},\mathcal{P})$ 上的复值函数

$$\xi(\omega)=\eta(\omega)+i\zeta(\omega)$$

为复随机变数，如果它的实、虚部分 $\eta(\omega),\zeta(\omega)$ 是两个 (实值) 随机变数． $\xi(\omega)$ 的数学期望 $E\xi$ 定义为

$$E\xi=E\eta+iE\zeta,$$

只要 $E\eta,E\zeta$ 存在．

今如对每一 $t\in T(\subset R_1)$, 存在一复随机变数 $\xi_t(\omega)$, 就称 $\{\xi_t(\omega),t\in T\}$ 为一复随机过程．

以后用 $\bar{a}$ 表 $a=b+ic$ 的共轭数，即 $\bar{a}=b-ic$.

设 $\{\xi_t(\omega),t\in R_1\}$ 为复随机过程，如果对一切 $t,\tau\in R_1$ 等式

$$B(\tau)\equiv E\xi_{t+\tau}\bar{\xi}_t=E\xi_\tau\bar{\xi}_0 \tag{10}$$

成立 (这里自然蕴含着一切 $E\xi_{t+\tau}\bar{\xi}_t$ 都存在的假定), 就称此过程为弱平稳过程；并称 $B(\tau)(\tau\in R_1)$ 为它的相关函数， $B(\tau)$ 一般取复值．

相关函数 $B(\tau)$ 是弱平稳过程的重要特征，它决定了过程的二阶矩性质．本节以下的内容就围绕 $B(\tau)$ 来叙述．

以上我们只讨论了过程的参数集 $T=R_1$ 的情形，对其它的 T 可类似地下定义，特别当 $T=N\equiv(0,\pm1,\pm2,\cdots)$ 时，我们得到了弱平稳 **序列** $\{\xi_n,n\in N\}$.

(二) **例**．

例 1 设 $\{\xi_n,n\in N\}$ 为复随机变数列，满足条件

$$E\xi_{n+\tau}\xi_n=\begin{cases}1, & \text{如}\tau=0,\\ 0, & \text{如}\tau\neq0.\end{cases} \tag{11}$$

这时， $B(\tau)=E\xi_{n+\tau}\bar{\xi}_n$ 与 n 无关，故是一弱平稳序列．

特别，如 $\{\xi_n,n\in N\}$ 取实值，相互独立，而且

$$E\xi_n=0,\quad D\xi_n=1, \tag{12}$$

则 (11) 满足．物理上称此 $\{\xi_n,n\in N\}$ 为"白噪声", 它是一种随机干扰的数学模型，这种干扰普遍存在于电流的波动、通信设备各部分的波动、电子发射的波动等各种波动现象中．

例 2 设 $\{z_n,n\geqslant1\}$ 为一列复随机变数，使

$$\begin{aligned}Ez_{n+\tau}\bar{z}_n&=b_n,\ (\tau=0),\\ &=0,\ (\tau\neq0),\end{aligned}$$

其中 $\sum\limits_{n=1}^{\infty}b_n<\infty$(注意 $b_n=E|z_n|^2$ 为实数), 任取一列实数 $\{\lambda_k,k\geqslant1\}$ 而定义

$$\xi_t=\sum_{k=1}^{\infty}z_ke^{i\lambda_kt},\ (t\in R_1) \tag{13}$$

这级数的收敛性用均方收敛 (§3.1). 我们证明，$\{\xi_t, t\in R_1\}$ 是弱平稳过程.

先证 (13) 中级数的确是均方收敛的. 为此，只要验证柯西 (Cauchy) 的收敛判别准则 (它对均方收敛仍有效) 满足. 当 $n>m$ 而 $m\to\infty$ 时

$$E\left|\sum_{k=m}^{n} z_k e^{i\lambda_k t}\right|^2=\sum_{k=m}^{n}E|z_k|^2=\sum_{k=m}^{n}b_k\longrightarrow 0.$$

为了验证弱平稳性，需要下列有用的事实:

$$E(\underset{n\to\infty}{\text{l.i.m}}\ \eta_n)=\lim_{n\to\infty}E\eta_n \tag{14}$$

这里 l.i.m 表均方收敛. 实际上，设 $\eta=\underset{n\to\infty}{\text{l.i.m}}\ \eta_n$, 由柯西 - 布尼亚科夫斯基不等式

$$|E\eta-E\eta_n|=|E(\eta-\eta_n)|\leqslant\sqrt{E|\eta-\eta_n|^2}\longrightarrow 0\,(n\to\infty).$$

再来验证弱平稳性: 由 (13)

$$\begin{aligned}E\xi_{t+\tau}\bar{\xi}_t&=E\left[\left(\sum_{k=1}^{\infty}z_k e^{i\lambda_k(t+\tau)}\right)\left(\sum_{j=1}^{\infty}\bar{z}_j e^{-i\lambda_j t}\right)\right]\\&=\sum_{k=1}^{\infty}E|z_k|^2e^{i\lambda_k\tau}=\sum_{k=1}^{\infty}b_k e^{i\lambda_k\tau}.\end{aligned}\tag{15}$$

故 $E\xi_{t+\tau}\bar{\xi}_t$ 与 t 无关. 对照 (13)(15) 是有益的.

例 3 考虑一个具有参数为 λ 的普阿松流 (§2.4), 以 τ_j 表此流中第 $j(j\geqslant 1)$ 个质点出现的时刻，因而以概率 1 有

$$0\equiv\tau_0<\tau_1<\tau_2<\cdots;$$

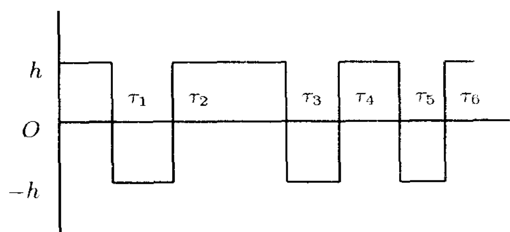


图 1

今定义

$$\begin{aligned}\xi_t&=h,\quad \text{如}\,t\in[\tau_{2j},\tau_{2j+1}),\\&=-h,\quad \text{如}\,t\in[\tau_{2j+1},\tau_{2j+2});\\&(j=0,1,2,\cdots)\end{aligned}\tag{16}$$

则 $\{\xi_t, t\geqslant 0\}$ 是一弱平稳过程.

实际上，由 (16),$\xi_{t+\tau}\xi_t$ 的值是 h^2 或 $-h^2$, 随区间 $(t,t+\tau],\tau>0$(或 $(t+\tau,t],\tau<0$) 中包含偶数个或奇数个 τ_j 而定，这里 0 算偶数. 由 §2.4, 知长为 $|\tau|$ 的区间 $(t,t+\tau]$ 或 $(t+\tau,t]$ 内含 k 个 τ_j 的概率为

$$p_k=e^{-\lambda|\tau|}\cdot\frac{\lambda^k|\tau|^k}{k!}. \tag{17}$$

故

$$\begin{aligned}E\xi_{t+\tau}\xi_t&=\sum_{k=0}^{\infty}h^2p_{2k}-\sum_{k=0}^{\infty}h^2p_{2k+1}\\&=h^2e^{-\lambda|\tau|}\sum_{k=0}^{\infty}\frac{(-\lambda|\tau|)^k}{k!}=h^2e^{-2\lambda|\tau|}\end{aligned}$$

与 t 无关，故 $\{\xi_t, t \geqslant 0\}$ 是弱平稳过程.

这例在电报信号中有用，信号由异号的电流符号发送，有一定的持续时间，电流值只取 h 或 $-h$, 我们假定此值改号的时刻 $\{\tau_j\}$ 构成一普阿松流 (见图 1).

例 4 实值随机过程 $\{\xi_t, t \in R_1\}$ 称为正态的 (或高斯 [Gauss] 的), 如果其中任意有穷多个 $\xi_{t_1}, \cdots, \xi_{t_n}$ 的联合分布是正态的. 设 $\xi_{t_1}, \cdots, \xi_{t_n}$ 的联合密度函数为

$$f_{\xi_{t_1},\cdots,\xi_{t_n}}(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{j,k=1}^{n} r_{jk}(x_j - a_j)(x_k - a_k)\right\}, \tag{18}$$

其中 B 为对称正定 n 级矩阵， $(r_{jk}) = B^{-1}$. 由 §2.12 定理 2, 知此过程的二阶矩存在，而且

$$\left.\begin{aligned} &E\xi_{t_j} = a_j, \\ &E(\xi_{t_j} - a_j)(\xi_{t_k} - a_k) = b_{jk}. \end{aligned}\right\} \tag{19}$$

今设

$$E\xi_t \equiv a\,(\text{常数}). \tag{20}$$

我们证明，对满足 (20) 的正态过程，平稳性与弱平稳性等价. 实际上，由正态性， $E\xi_{t+\tau}\xi_t$ 存在，据平稳性知它与 t 无关，故得弱平稳性. 反之，由正态性知 $\xi_{t_1}, \cdots, \xi_{t_n}$ 的密度为 (18), 又 $\xi_{t_1+\tau}, \cdots, \xi_{t_n+\tau}$ 的密度为

$$f_{\xi_{t_1+\tau},\cdots,\xi_{t_n+\tau}}(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\widetilde{B}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\sum_{j,k=1}^{n} \widetilde{r}_{jk}(x_j - \widetilde{a}_j)(x_k - \widetilde{a}_k)\right\},$$

由 (20) 及 (19) 及弱平稳性知

$$a_j = E\xi_{t_j} = a = E\xi_{t_j+\tau} = \widetilde{a}_j,$$

$$\begin{aligned} b_{jk} &= E(\xi_{t_j} - a)(\xi_{t_k} - a) = E\xi_{t_j}\xi_{t_k} - a^2 \\ &= E\xi_{t_j+\tau}\xi_{t_k+\tau} - a^2 = E(\xi_{t_j+\tau} - a)(\xi_{t_k+\tau} - a) = \widetilde{b}_{jk}. \end{aligned}$$

于是

$$f_{\xi_{t_1},\cdots,\xi_{t_n}}(x_1, \cdots, x_n) = f_{\xi_{t_1+\tau},\cdots,\xi_{t_n+\tau}}(x_1, \cdots, x_n),$$

故过程是平稳的.

(三) **相关函数**. 设 $B(\tau)(\tau \in R_1)$ 是弱平稳过程 $\{\xi_t, t \in R_1\}$ 的相关函数，它必定具有下列性质：

1) $B(0) \geqslant 0$;

2) $|B(\tau)| \leqslant B(0)$;

3) $B(-\tau) = B(\tau)$;

4) 非负定性；对任意正整数 n, 任意实数 t_i, 任意复数 $a_i, i = 1, 2, \cdots, n$, 有

$$\sum_{j,k=1}^{n} B(t_j - t_k)a_j\overline{a}_k \geqslant 0.$$

实际上， $B(0)=E|\xi_t|^2\geqslant 0$; 又

$$|B(\tau)|^2=|E\xi_{t+\tau}\overline{\xi}_t|^2\leqslant E|\xi_{t+\tau}|^2\cdot E|\xi_t|^2=[B(0)]^2,$$

$$B(-\tau)=E\overline{\xi}_{t+\tau}\xi_t=E\overline{\xi_{t+\tau}\overline{\xi}_t}=\overline{B(\tau)}.$$

最后

$$\sum_{j,k=1}^{n}B(t_j-t_k)a_j\overline{a}_k=E\Big[\sum_{j,k=1}^{n}\xi_{t_j}\overline{\xi}_{t_k}a_j\overline{a}_k\Big]$$
$$=E\Big[\Big(\sum_{j=1}^{n}\xi_{t_j}a_j\Big)\overline{\Big(\sum_{k=1}^{n}\xi_{t_k}a_k\Big)}\Big]=E|\sum_{j=1}^{n}\xi_{t_j}a_j|^2\geqslant 0.$$

重要的一类弱平稳过程是均方连续的弱平稳过程，称弱平稳过程 $\{\xi_t,t\in R_1\}$ 为均方连续的，如果对任意的 $t\in R_1$, 有

$$\lim_{h\to\infty}E|\xi_{t+h}-\xi_t|^2=0.$$

我们所以考虑 **均方** 收敛下的连续性是因为均方收敛是二阶矩性质.

引理 1 弱平稳过程 $\{\xi_t,t\in R_1\}$ 均方连续的充要条件是它的相关函数 $B(\tau)$ 在 R_1 上连续.

证 由等式

$$E|\xi_{t+\tau}-\xi_t|^2=E[\xi_{t+\tau}-\xi_t]\overline{[\xi_{t+\tau}-\xi_t]}=2B(0)-B(\tau)-\overline{B(\tau)}. \tag{21}$$

得证充分性. 必要性则由下式得到:

$$|B(t+\tau)-B(t)|=|E\xi_{t+\tau}\overline{\xi}_0-E\xi_t\overline{\xi}_0|=|E(\xi_{t+\tau}-\xi_t)\overline{\xi}_0|$$
$$\leqslant\{E|\xi_{t+\tau}-\xi_t|^2\cdot E|\xi_0|^2\}^{\frac{1}{2}}=\{E|\xi_{t+\tau}-\xi_t|^2\cdot B(0)\}^{\frac{1}{2}}.\ \square$$

由此得到下列重要的相关函数的谱分解定理:

定理 1(辛钦 Хинчин) 如 $B(\tau)$ 是均方连续弱平稳过程的相关函数，则

$$B(\tau)=\int_{-\infty}^{\infty}e^{i\tau x}dF(x), \tag{22}$$

其中 $F(x)$ 是 R_1 上的单调不减，右连续函数，而且

$$\lim_{x\to-\infty}F(x)=0,\quad \lim_{x\to\infty}F(x)=B(0). \tag{23}$$

证 设 $B(0)>0$. 根据引理 1 及性质 4), 知 $\frac{B(\tau)}{B(0)}$ 是连续的非负定函数，由 §2.13 波赫纳尔 - 辛钦定理，知存在分布函数 $G(x)$, 满足

$$\frac{B(\tau)}{B(0)}=\int_{-\infty}^{\infty}e^{i\tau x}dG(x).$$

令 $F(x)=B(0)G(x)$, 即知 (22)(23) 成立.

如 $B(0)=0$, 由性质 2),$B(\tau)\equiv 0$, 取 $F(x)\equiv 0$ 即可. □

由定理 1 及波赫纳尔 - 辛钦定理可见，均方连续弱平稳过程的相关函数 $B(\tau)$ 如不恒等于 0, 则除一常数因子外，重合于一特征函数.

(22) 式为相关函数的谱展式，其中的 $F(x)$ 称为过程的谱函数. 如果存在非负函数 $f(x),(x\in R_1)$, 使对任意 $x\in R_1$ 有

$$F(x)=\int_{-\infty}^{x} f(y)dy, \tag{24}$$

则称 $f(x)$ 为过程的谱密度，这时由 (22) 得

$$B(\tau)=\int_{-\infty}^{\infty} e^{i\tau x}f(x)dx. \tag{25}$$

当 $\int_{-\infty}^{\infty}|B(\tau)|d\tau<-\infty$ 时，根据 §2.11 系 2, 有

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ix\tau}B(\tau)d\tau. \tag{26}$$

当 $B(\tau)$ 取实值时，由性质 2),$B(\tau)=B(-\tau)$ 又因 $e^{-ix\tau}=\cos x\tau-i\sin x\tau$ ，而且 $f(x)$ 取实值，故得

$$f(x)=\frac{1}{\pi}\int_{0}^{\infty} B(\tau)\cos\tau x d\tau. \tag{27}$$

类似地，对任意弱平稳序列 $\{\xi_n,n\in N\}$ 的相关函数 $B(n),(n\in N)$, 利用 §3.7 赫尔格洛兹定理，得

$$B(n)=\int_{-\pi}^{\pi} e^{inx}dF(x),$$

其中 $F(x)$ 是 $[-\pi,\pi]$ 上的单调不减右连续函数，满足

$$F(-\pi)=0,\quad F(\pi)=B(0), \tag{28}$$

例 2′ 继续讨论例 2, 已知 (13) 中过程的相关函数为

$$B(\tau)=\sum_{k=1}^{\infty} b_k e^{i\lambda_k\tau},\ (\tau\in R_1). \tag{29}$$

因为 $|b_k e^{i\lambda_k\tau}|\leqslant b_k$, $\sum\limits_{k=0}^{\infty} b_k<\infty$, 故 $\sum\limits_{k=1}^{\infty} b_k e^{i\lambda_k\tau}$. 关于 τ 均匀收敛，从而可在求和号下取极限而得

$$\lim_{s\to\tau} B(s)=B(\tau),$$

这说明 $B(\tau)$ 连续. 由引理 1 知过程均方连续. 定义

$$F(x)=\sum_{k:\lambda_k\leqslant x} b_k, \tag{30}$$

其中求和号对一切满足 $\lambda_k \leqslant x$ 的 k 进行，则可把 (29) 改写为

$$B(\tau)=\int_{-\infty}^{\infty} e^{i\tau x}dF(x).$$

这就是相关函数的谱展式，(30) 中的 $F(x)$ 为谱函数.

例 3′ 由例 3 知 (16) 中过程的相关函数为

$$B(\tau)=h^2e^{-2\lambda|\tau|},$$

这是一连续函数，故过程均方连续. 由于

$$\int_{-\infty}^{\infty}|B(\tau)|d\tau=h^2\int_{-\infty}^{\infty}e^{-2\lambda|\tau|}d\tau=2h^2\int_0^{\infty}e^{-2\lambda\tau}d\tau<\infty$$

由 (27) 得谱密度为

$$f(x)=\frac{1}{\pi}\int_0^{\infty}h^2e^{-2\lambda\tau}\cos\tau x d\tau=\frac{2}{\pi}\cdot\frac{h^2\lambda}{x^2+4\lambda^2}.$$

若干相关函数与谱密度对应表

相关函数 $B(\tau)$	谱密度 $f(x)$
1. $\sigma^2e^{-\alpha\|\tau\|}\quad(\alpha>0)$	1. $\frac{\sigma^2\alpha}{\pi(\alpha^2+x^2)}$
2. $\sigma^2e^{-\alpha\tau^2}\quad(\alpha>0)$	2. $\frac{\sigma^2}{2\sqrt{\pi\alpha}}\exp\left(-\frac{x^2}{4\alpha}\right)$
3. $\sigma^2e^{-\alpha\tau^2}\cos\beta\tau\quad(\alpha>0)$	3. $\frac{\sigma^2}{4\sqrt{\pi\alpha}}\left\{\exp\left[-\frac{(x+\beta)^2}{4\alpha}\right]+\exp\left[-\frac{(x-\beta)^2}{4\alpha}\right]\right\}$
4. $\sigma^2e^{-\alpha\|\tau\|}\cos\beta\tau\quad(\alpha>0)$	4. $\frac{\sigma^2\alpha}{2\pi}\left[\frac{1}{(x-\beta)^2+\alpha^2}+\frac{1}{(x+\beta)^2+\alpha^2}\right]$
5. $\sigma^2e^{-\alpha\|\tau\|}\left(\cos\beta\tau+\frac{\alpha}{\beta}\sin\beta\|\tau\|\right),(\alpha>0,\beta>0)$	5. $\frac{2\sigma^2}{\pi}\frac{\alpha(\alpha^2+\beta^2)}{(x^2+\alpha^2-\beta^2)^2+4\alpha^2\beta^2}$
6. $2a\frac{\sin b\tau}{\tau}$	6. $f(x)=\begin{cases}a, & \|x\|\leqslant b,\\ 0, & \|x\|>b.\end{cases}$
7. $2a^2(2\cos b\tau-1)\frac{\sin b\tau}{\tau}$	7. $f(x)=\begin{cases}0 & \|x\|<b,\\ a^2 & b\leqslant\|x\|\leqslant 2b,\\ 0 & 2b<\|x\|.\end{cases}$
8. $Ae^{-\alpha\|\tau\|}(1+\alpha\|\tau\|)$	8. $\frac{2A\alpha^3}{\pi(x^2+\alpha^2)^2}$
9. $B(\tau)=\begin{cases}\sigma^2(1-\|\tau\|), & \|\tau\|\leqslant 1,\\ 0, & \|\tau\|>1.\end{cases}$	9. $\frac{\sigma^2}{2\pi}\left(\sin\frac{x}{2}\Big/\frac{x}{2}\right)^2$
10. $Ae^{-\alpha\|\tau\|}(1+\alpha\|\tau\|+\frac{1}{3}\alpha^2\tau^2)$	10. $\frac{8A\alpha^5}{3\pi(x^2+\alpha^2)^3}$
11. $Ae^{-\alpha\|\tau\|}\left(\cos\beta\tau-\frac{\alpha}{\beta}\sin\beta\|\tau\|\right)$	11. $\frac{2A\alpha x^2}{\pi\left[(x^2+\alpha^2+\beta^2)^2-4\beta^2x^2\right]}$

作为随机过程论的初步介绍就到此为止；附带补充几句话.

随机过程论是概率论的一个重要分支，近三、四十年来，这门数学学科有了蓬勃的发展，积累了非常多的资料. 我们在本章内，对它只是作了一个初步的介绍. 读者如有志于此，可以看专门的书：关于过程论的较全面的叙述见参考书 [2],[16],[17],[18], 侧重于应用的书见 [19],[20],[21]. 关于平稳过程的专著见 [22],[23],[24],[25]; 关于马尔科夫过程的专著见 [26],[27],[28],[21]; 有关马尔科夫链的深入论述见 [29]; 关于这两类过程的基本理论见 [35].

习　　题

1. 设齐次马氏链的转移概率矩阵为

$$P=\begin{pmatrix}1/2 & 1/3 & 1/6\\ 1/3 & 1/3 & 1/3\\ 1/3 & 1/2 & 1/6\end{pmatrix}$$

问此链共有几个状态？求二步转移矩阵. $\lim\limits_{n\to\infty} p_{jk}^{(n)}$ 是否存在？并求之. 此链是否遍历？

2. 从 1,2,3,4,5,6 六个数中，等可能地取出一数，取后还原，不断独立地连取下去，如果在前 n 次中所取得的最大数是 j, 就说质点在第 n 步时位置在状态 j. 这质点的运动构成一齐次马氏链，试写出状态空间及转移概率矩阵.

3. 设 $\{\xi_t, t\in[0,\infty)\}$ 是普阿松过程， $\xi_0\equiv 0$, 试求它的有穷维分布函数族.

4. 设 $\{\xi_t, t\in[0,\infty)\}$ 是维纳过程， $\xi_0\equiv 0$, 试求它的有穷维分布函数族. （设 $\sigma=1$）

5. 设 $\{\xi_t, t\in R_1\}$ 是均方连续弱平稳过程，具有相关函数 $B_\xi(\tau)$ 及谱函数 $F_\xi(x)$, 取复常数 a_k 及实常数 $s_k, k=1,\cdots,n$, 并定义

$$\eta_t=\sum_{k=1}^{n} a_k\xi_{t+s_k},\ (t\in R_1).$$

试证 $\{\eta_t, t\in R_1\}$ 也是均方连续弱平稳过程，并求这过程的相关函数 $B_\eta(\tau)$ 及谱函数 $F_\eta(x)$.

第五章　数理统计初步

§5.1　基本概念

(一) 引言. 到现在为止，我们总是从已给的随机变数 ξ 出发，来讨论 ξ 的种种性质，这时 ξ 的分布函数 $F(x)$ 都已事先给定. 然而在实际问题中， $F(x)$ 常常是未知的.

我们来考虑一个具体的例子.

某砖厂生产青砖，由于原料和生产过程的种种随机因素，各块砖的抗压强度一般是不同的. 以 ω 来记一块砖， $\xi(\omega)$ 表这块砖的抗压强度，当 ω 变动时，我们便得到一随机变数. 试问如何求出 (或近似地求出)ξ 的分布函数

$$F(x)=p\,(\xi\leqslant x)?$$

基于长期的生产经验，人们提出假设：这分布是 $N(a,\sigma)$ 正态的. 试问如何来判断这假设的正确性？在某些实际问题里，有时我们并不需要完全确定 $F(x)$, 而只需要知道砖的平均抗压强度 $E\xi$ 或 ξ 的其它数字特征，试问如何来估计 $E\xi$?

关于最后一个问题，通常会这样想：由于全部青砖所成的母体的块数太多，不可能一一计算，我们只好从其中随机地取出几块而观察它们的抗压强度 $\xi_1,\cdots,\xi_n$, 这 n 块砖构成一容量为 n 的子样，然后以这子样的平均值 $\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 作为母体的数学期望.

应该从两种观点来考虑子样 $(\xi_1,\cdots,\xi_n)$ 的函数

$$g(\xi_1,\cdots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i. \tag{1}$$

一方面，在抽样以前，每个 ξ_i 的值可以是 ξ 所能取的值中的任一个，我们不能准确地预言它的值，因而生产条件不变时，每个 ξ_i 可看成为与 ξ 有相同分布的随机变数；如果每次取样是还原地独立进行的，还可假定 $\xi_1,\cdots,\xi_n$ 相互独立；这时，作为随机变数 $(\xi_1,\cdots,\xi_n)$ 的函数， $g(\xi_1,\cdots,\xi_n)$ 也是一随机变数. 另方面，在抽样以后， ξ_i 的值已完全确定是一常数，这常数是对 ξ_i 的一次观察值，不过我们仍记它为 ξ_i, 因而 $g(\xi_1,\cdots,\xi_n)$ 也是一常数 [1].

本章中，我们只考虑数理统计中 **一些** 基本问题，设已给代表母体的随机变数 ξ, 要研究的是：

1° 未知分布函数的估计：试根据对 ξ 的 n 个独立的观察 $\xi_1,\cdots,\xi_n$ 来估计 ξ 的分布函数，有时也称 $(\xi_1,\cdots,\xi_n)$ 的容量为 n 的子样.

2° 未知参数的估计：设 a 为 ξ 的某数字特征 (例如数学期望、方差等等), 试根据子样 $\xi_1,\cdots,\xi_n$ 来估计 a 的值.

如果取一个函数 $g(\xi_1,\cdots,\xi_n)$ 来做为 a 的估值，便称 $g(\xi_1,\cdots,\xi_n)$ 为 a 的点估值. 如果取两个函数 $g_1=g_1(\xi_1,\cdots,\xi_n)$ 及 $g_2(\xi_1,\cdots,\xi_n)$, 使以充分大的概率，不等式

$$g_1<a<g_2$$

---

[1] 数学上严格地说，不固定 ω 时， $g(\xi_1(\omega),\cdots,\xi_n(\omega))$ 是一随机变数；但固定 ω 时， $g(\xi_1(\omega),\cdots,\xi_n(\omega))$ 是一常数. 每抽完 n 个样，相当于固定一次 $\omega,\omega=(\omega_1,\cdots,\omega_n)$ 是复合随机试验 $\widetilde{E}=E^n$ 的一个基本事件， E 表抽样一次的试验.

成立，亦即区间 (g_1, g_2) 包含常数 a, 那么便称这种估值方法为区间估值.

3° 统计假设的检验：根据子样 $\xi_1, \cdots, \xi_n$, 及预先给定的概率，来肯定或否定某一假设 H. H 可以是，譬如说：

H : 未知分布是某分布 (例如，是正态分布，普阿松分布等等).

H : 未知的参数取某定值或属于某集合.

我们注意，以上问题远未穷尽数理统计中的全部基本问题.

(二) **未知分布函数的估计.**

设 ξ 为随机变数，它的分布函数 $F(x)$ 未知，对 ξ 进行 n 次独立的观察而得

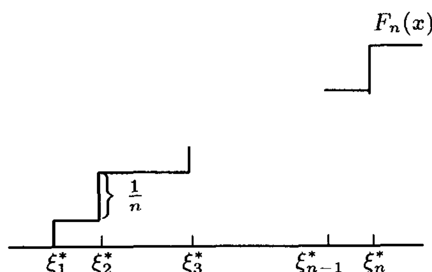
$$\xi_1, \xi_2, \cdots, \xi_n, \tag{2}$$

把它们按大小排列为

$$\xi_1^* \leqslant \xi_2^* \leqslant \cdots \leqslant \xi_n^* \tag{3}$$

这种按单调不减排列后的序列很是重要. 定义函数 $F_n(x)$:

$$F_n(x) = \begin{cases} 0, & x < \xi_1^* \text{ 时}, \\ \dfrac{k}{n}, & \xi_k^* \leqslant x < \xi_{k+1}^* \text{ 时}, \\ 1, & x \geqslant \xi_n^* \text{ 时}. \end{cases}$$



当 $\xi_1^*, \cdots, \xi_n^*$ 的值固定时，$F_n(x)$ 是一分布函数，只能在 $\xi_k^* (k = 1, \cdots, n)$ 处有断点，跃度是 $\frac{1}{n}$ 的倍数 (如有 l 个元相重，$\xi_{k-1}^* < \xi_k^* = \xi_{k+1}^* = \cdots = \xi_{k+l-1}^* < \xi_{k+l}^*$, 则在断点 ξ_k^* 上的跃度为 $\frac{l}{n}$).

称 $F_n(x)$ 为 ξ 的经验分布函数.

现在采用另一观点来考虑 $F_n(x)$, **这是下面一切讨论的基本观点**. 如前所述，(2) 中的 $\xi_1, \cdots, \xi_n$ 可看成是 n 个独立的具有相同分布函数 $F(x)$ 的随机变数，因而对每个固定的 $x \in R_1$, 作为 $\xi_1, \cdots, \xi_n$ 的函数，$F_n(x)$ 也是一随机变数. 实际上

$$P\left(F_n(x) = \frac{k}{n}\right) = P(\text{ 恰有} k \text{ 个} \xi_i : \xi_i \leqslant x). \tag{4}$$

为了计算右方值，注意 n 次独立观察可看成 n 重贝努里试验 $\widetilde{E} = E^n$, 其中 E 为一次观察. 以 A 表事件 "观察值落于 $(-\infty, x]$ 中", 则上式右方概率等于 $\widetilde{E}$ 中 A 出现 k 次的概率，因为 ξ_i 与 ξ 有相同的分布函数，故每次试验中 A 出现的概率为 $F(x)$, 从而

$$P\left(F_n(x) = \frac{k}{n}\right) = \binom{n}{k} \left[F(x)\right]^k \left[1 - F(x)\right]^{n-k}. \tag{5}$$

下面的格里文科 (Гливенко 定理指出，当 $n \to \infty$ 时，以概率 $1, \{F_n(x)\}$ 关于 x 均匀地趋于 $F(x)$. 作为准备，先述

引理 1 设 $\{A_k\}, k = 1, 2, \cdots$ 是一事件列，如果

$$P(A_k) = 1, \ (k = 1, 2, \cdots) \tag{6}$$

则

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1.$$

证 由 §1.3(7′) 及 §1.3(11) 得

$$P\left(\overline{\bigcap_{k=1}^{\infty} A_k}\right) = P\left(\bigcup_{k=1}^{\infty} \overline{A}_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} \overline{A}_k\right)$$
$$\leqslant \lim_{n\to\infty} \sum_{k=1}^{n} P(\overline{A}_k) = 0. \quad \square$$

定理 1 (格里文科) 设 $F(x)$ 是随机变数 ξ 的分布函数, $F_n(x)$ 是 ξ 的经验分布函数, 则当 $n \to \infty$ 时有

$$P\left(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\right) = 1. \tag{7}$$

证 简记 $D_n = \sup\limits_{-\infty<x<\infty} |F_n(x) - F(x)|$. 由 $F_n(x), F(x)$ 的右连续性, 得 $D_n = \sup\limits_{x\in M} |F_n(x) - F(x)|$, 其中 M 是全体有理数集, 故 M 可列. 对 $a \in R_1$,

$$(D_n \leqslant a) = \bigcap_{x\in M} \big(|F_n(x) - F(x)| \leqslant a\big),$$

作为可列多个事件的交, $(D_n \leqslant a)$ 也是一事件, 因而 D_n 是随机变数. 根据 §3.1(10) 式, 知 $(\lim\limits_{n\to\infty} D_n = 0)$ 是一事件而 (7) 中左方 $P(\lim\limits_{n\to\infty} D_n = 0)$ 有意义.

对任意整数 r, 令

$$x_{r,k} = \inf\left(x : F(x-0) \leqslant \frac{k}{r} \leqslant F(x) = F(x+0)\right), k = 1, 2, \cdots, r.$$

以 A 表事件 "$\xi < x_{r,k}$", 显然 $P(A) = F(x_{r,k} - 0)$. 因为在 n 次独立观察中, 事件 A 出现的频率为 $F_n(x_{r,k} - 0)$, 故由强大数定理 (§3.3 系 4) 得

$$P\left(\lim_{n\to\infty} F_n(x_{r,k} - 0) = F(x_{r,k} - 0)\right) = 1. \tag{8}$$

类似地, 考虑事件 "$\xi \leqslant x_{r,k}$" 后, 可得

$$P\left(\lim_{n\to\infty} F_n(x_{r,k}) = F(x_{r,k})\right) = 1. \tag{9}$$

定义事件

$$B_k^r = \left(\lim_{n\to\infty} F_n(x_{r,k} - 0) = F(x_{r,k} - 0)\right) \cap \left(\lim_{n\to\infty} F_n(x_{r,k}) = F(x_{r,k})\right),$$

$$B^r = \bigcap_{k=1}^{r} B_k^r.$$

由 (8)(9) 及引理 1, 得 $P(B_k^r)=1, P(B^r)=1$. 易见

$$B^r=\Big(\lim_{n\to\infty}\max_{1\leqslant k\leqslant r}|F_n(x_{r,k}-0)-F(x_{r,k}-0)|=0\Big)$$
$$\cap\Big(\lim_{n\to\infty}\max_{1\leqslant k\leqslant r}|F_n(x_{r,k})-F(x_{r,k})|=0\Big). \tag{10}$$

再令 $B=\bigcap_{r=1}^{\infty}B^r$, 仍由引理 1, 有 $P(B)=1$.

对任意 $x\in[x_{r,k},x_{r,k+1})$, 有

$$F_n(x_{r,k})\leqslant F_n(x)\leqslant F_n(x_{r,k+1}-0),$$

$$F(x_{r,k})\leqslant F(x)\leqslant F(x_{r,k+1}-0),$$

从而

$$F_n(x_{r,k})-F(x_{r,k+1}-0)\leqslant F_n(x)-F(x)\leqslant F_n(x_{r,k+1}-0)-F(x_{r,k}). \tag{11}$$

根据

$$0\leqslant F(x_{r,k+1}-0)-F(x_{r,k})\leqslant\frac{1}{r}, \tag{12}$$

可以证明

$$\sup_{x_{r,1}\leqslant x\leqslant x_{r,r}}|F_n(x)-F(x)|\leqslant\max_{1\leqslant k\leqslant r}|F_n(x_{r,k})-F(x_{r,k})|$$
$$+\max_{1\leqslant k\leqslant r}|F_n(x_{r,k}-0)-F(x_{r,k}-0)|+\frac{1}{r}. \tag{13}$$

实际上，如 $F_n(x)-F(x)\geqslant 0$, 由 (11)(12) 得

$$F_n(x)-F(x)\leqslant|F_n(x_{r,k+1}-0)-F(x_{r,k+1}-0)|+|F(x_{r,k+1}-0)-F(x_{r,k})|$$
$$\leqslant\max_{1\leqslant k\leqslant r}|F_n(x_{r,k}-0)-F(x_{r,k}-0)|+\frac{1}{r}; \tag{14}$$

如 $F_n(x)-F(x)\leqslant 0$, 类似有

$$|F_n(x)-F(x)|\leqslant|F(x_{r,k+1}-0)-F(x_{r,k})|+|F(x_{r,k})-F_n(x_{r,k})|$$
$$\leqslant\max_{1\leqslant k\leqslant r}|F_n(x_{r,k})-F(x_{r,k})|+\frac{1}{r}. \tag{15}$$

由于 (14),(15) 对任意 $x\in[x_{r,k},x_r,x_{r,k+1})$ 正确，故 (13) 得证.

由 (13) 及 (10) 可见

$$B^r\subset\Big(\lim_{n\to\infty}\sup_{x_{r,1}\leqslant x\leqslant x_{r,r}}|F_n(x)-F(x)|\leqslant\frac{1}{r}\Big). \tag{16}$$

注意，$x_{r,r}$ 是一不依赖于 r 的常数，记为 b, 而 $x_{r,1}$ 是 r 的不增函数，故 $x_{r,1}\downarrow a, -\infty\leqslant a<b\leqslant\infty$, 当 a 或 b 有穷时，如 $x<a$, 则 $F_n(x)=F(x)=0$; 如 $x\geqslant b$, 则 $F_n(x)=F(x)=1$, 因而由 (16) 得

$$B\subset\Big(\lim_{n\to\infty}D_n\leqslant\frac{1}{k}\Big),$$

其中 k 为任一正整数，由 k 的任意性

$$B \subset \left(\lim_{n\to\infty} D_n = 0\right).$$

既然 $P\left(\lim\limits_{n\to\infty} D_n = 0\right)$ 有意义，我们得到

$$P\left(\lim_{n\to\infty} D_n = 0\right) \geqslant P(B) = 1.$$

于是 (7) 式获证. □

注意，如 (7) 中的 sup 去掉，那么，对每个固定的 $x \in R_1$ 有

$$\lim_{n\to\infty} |F_n(x) - F(x)| = 0, \quad \text{(a.s.)}$$

由强大数定理，这是显然成立的 [1], 故格里文科定理的主要点正在于此 sup, 即上式中收敛对 x 的均匀性.

(三) **子样的数字特征**. 经验分布函数 $F_n(x)$ 是离散型的，这分布函数的密度矩阵是

$$\begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_n \\ \dfrac{1}{n} & \dfrac{1}{n} & \cdots & \dfrac{1}{n}. \end{pmatrix} \tag{17}$$

当观察值 $\xi_1, \cdots, \xi_n$ 固定时，它是一普通的分布函数，它的数字特征容易求出，例如

$$\text{平均值 } \overline{\xi} = \frac{1}{n}\sum_{j=1}^{n} \xi_j, \tag{18}$$

$$\text{方差 } s^2 = \frac{1}{n}\sum_{j=1}^{n} (\xi_j - \overline{\xi})^2, \tag{19}$$

等等，我们分别称 $\overline{\xi}, s^2$ 为子样平均值及子样方差. 类似地可以定义子样的任一数字特征为 $F_n(x)$ 的对应的数字特征，其中特别重要的是子样的 k 阶矩 (或中心矩), 中位数及众数等等.

值得注意的是：当子样变动时，$\xi_1, \cdots, \xi_n$ 是随机变数. 既然子样的数字特征由随机变数 $\xi_1, \cdots, \xi_n$ 决定，故这些数字特征本身也是随机变数.

由于经验分布函数 $F_n(x)$ 以概率 1 均匀趋向于 ξ 的分布函数 $F(x)$, 自然地想到：是否可以用子样的数字特征来作为 ξ 的对应的数字特征的估值？例如，当 n 甚大时，是否可用 (18) 中的 $\overline{\xi}$ 来近似地估计 $E\xi$? 这问题决定于我们对估值的要求，留待以后详细讨论.

(四) **直方图**. 关于分布函数 $F(x)$ 的近似求法已由格里文科定理解决，现在来讨论分布密度的近似计算问题，并介绍实际中常用的直方图.

先设 ξ 是离散型的，密度矩阵为

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}, \tag{20}$$

---

[1] 实际上，令 $\zeta_i = 1$, 如 $\xi_i \leqslant x$; $\zeta_i = 0$, 如 $\xi_i > x$ 则 $P(\zeta_i = 1) = F(x)$, 而且 $F_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} \zeta_i \to F(x)$, (a.s.).

其中 $p_i, i = 1, 2, \cdots$, 未知. 对 ξ 作 n 次独立的观察而得 $\xi_1, \cdots, \xi_n$, 令 f_i(依赖于 n) 为满足 $\xi_j = a_i$ 的 j 的个数，亦即事件 "$\xi = a_i$" 在此 n 次独立观察中的出现次数，则由强大数定理

$$\lim_{n\to\infty} \frac{f_i}{n} = p_i, (i = 1, 2, \cdots), \ (a.s.). \tag{21}$$

由此可见，当 n 充分大后，下面的图 1 与图 2 应很相似 (其中 ξ 只取 4 个值，$\sum_{i=1}^{4} p_i = 1$).

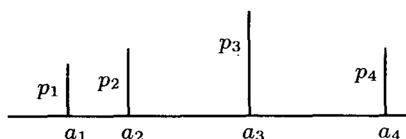


图 1

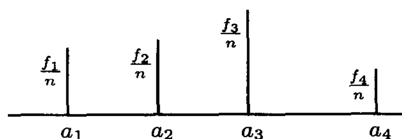


图 2

次设 ξ 是连续的，密度函数 $f(x)$ 未知. 对任一有限区间 $(a, b]$, 用下列分点分成 m 个子区间 $(m < n)$, 其长度不一定要相等

$$a = x_0 < x_1 < \cdots < x_{m-1} < x_m = b$$

对 ξ 的 n 次独立观察 $\xi_1, \cdots, \xi_n$ 中，落于 $(x_i, x_{i+1}]$ 中的设为 f_i 个，因而事件 "$x_i < \xi \leqslant x_{i+1}$" 的频率为 $\frac{f_i}{n}$. 由强大数定理

$$\lim_{n\to\infty} \frac{f_i}{n} = P(x_i < \xi \leqslant x_{i+1}) = \int_{x_i}^{x_{i+1}} f(x)dx, \ (a.s.).$$

因此，如 $f(x)$ 连续，则当 n 充分大时，有

$$\frac{f_i}{n} \approx \Delta x_i f(x_i), \ (\Delta x_i = x_{i+1} - x_i),$$

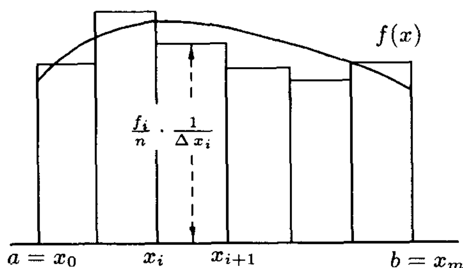


图 3

亦即

$$\frac{f_i}{n} \cdot \frac{1}{\Delta x_i} \approx f(x_i).$$

今定义函数

$$\varphi_n(x) = \frac{f_i}{n} \cdot \frac{1}{\Delta x_i},$$

当 $x_i < x \leqslant x_{i+1}$, 称 $\varphi_n(x)$, $x \in (a, b]$ 的图形为 $(a, b]$ 上的直方图. 当 n 及 m 充分大时，在 $(a, b]$ 上，这图近似于 $f(x)$ 的图 (见图 3).

§5.2 子样数字特征的分布

(一) 子样的平均值与方差. 设随机变数 ξ 的分布函数为 $F(x)$, 对 ξ 进行 n 次独立的观察而得一容量为 n 的子样

$$(\xi_1, \cdots, \xi_n). \tag{1}$$

前面已经讲过，可以假定 $\xi_1,\cdots,\xi_n$ 为独立同分布的随机变数，有相同的分布函数为 $F(x)$.

考虑子样的一个数字特征

$$G=g(\xi_1,\cdots,\xi_n),\tag{2}$$

其中 $g(x_1,\cdots,x_n)$, $(x_i\in R_1)$ 是某一 n 元波雷耳可测函数，因而 G 也是一随机变数 (§2.8(一)). 在数理统计的许多问题中，重要的是

1. 求出 G 的分布；关于求此分布的一般方法已在 §2.8 中讨论过.

2. 求出 G 的数字特征，如平均值，方差等.

3. 当 $g(\xi_1,\cdots,\xi_n)$ 的精确分布难以求出时，试求当 $n\to\infty$ 时，$g(\xi_1,\cdots,\xi_n)$ 的极限分布.

现在来举一些例子，以下假定

$$E\xi=a,\ D\xi=\sigma^2.$$

例 1 考虑子样平均值

$$\overline{\xi}=\frac{1}{n}\sum_{j=1}^{n}\xi_j.\tag{3}$$

注意 $\overline{\xi}$ 依赖于 n, 由于 ξ_j 与 ξ 有相同分布，故

$$E\overline{\xi}=\frac{1}{n}\sum_{j=1}^{n}E\xi_j=a,\tag{4}$$

$$D\overline{\xi}=\frac{1}{n^2}\sum_{j=1}^{n}D\xi_j=\frac{\sigma^2}{n},\tag{5}$$

这里出现了一重要事实：虽然 $E\overline{\xi}=E\xi$ 与 n 无关，但 $\overline{\xi}$ 的方差却只是 $D\xi$ 的 $\frac{1}{n}$.

设 ξ 有正态分布 $N(a,\sigma)$, 由 $\xi_1,\cdots,\xi_n$ 的独立性及正态分布的性质 (§2.11,(44)), 可见 $\overline{\xi}$ 也有正态分布为 $N(a,\frac{\sigma}{\sqrt{n}})$.

如果 ξ 不一定有正态分布，但方差非 0 而且有穷，根据中心极限定理 (§3.4, 定理 1), 得

$$P\left(\frac{\sum\limits_{j=1}^{n}\xi_j-na}{\sigma\sqrt{n}}\leqslant x\right)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}dz,\tag{6}$$

亦即

$$P\left(\frac{\overline{\xi}-a}{\frac{\sigma}{\sqrt{n}}}\leqslant x\right)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}dz,\ \ (n\to\infty).\tag{7}$$

为了方便，我们引进一定义：设对随机变数列 $\{Y_n\}$ 存在二列常数，$\{a_n\}$ 及 $\{\sigma_n\},\sigma_n\neq 0$ 使 $\frac{Y_n-a_n}{\sigma_n}$ 的分布函数 $F_n(x)$ 当 $n\to\infty$ 时趋于某分布函数 $F(x)$, 就说 Y_n 为 (a_n,σ_n) 渐近 $F(x)$ 的，特别，如 $F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}dz$, 就说 Y_n 为 (a_n,σ_n) 渐近正态的.

(7) 式表示，$\overline{\xi}$ 为 $(a,\frac{\sigma}{\sqrt{n}})$ 渐近正态的.

例 2 考虑子样的方差

$$s^2=\frac{1}{n}\sum_{j=1}^{n}(\xi_j-\overline{\xi})^2=\frac{1}{n}\sum_{j=1}^{n}(\xi_j-a)^2-(\overline{\xi}-a)^2. \tag{8}$$

由于 $E(\xi_j-a)^2=\sigma^2, E(\overline{\xi}-a)^2=D\overline{\xi}=\frac{\sigma^2}{n}$, 故

$$Es^2=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2 \tag{9}$$

与 (4) 式不同，这里子样方差的平均值 Es^2 不等于 ξ 的方差 σ^2, 尽管当 n 甚大时， Es^2 与 σ^2 相差甚微. 对固定的 n, 有时考虑 **修正** 子样方差

$$s^{*2}=\frac{n}{n-1}s^2=\frac{1}{n-1}\sum_{j=1}^{n}(\xi_j-\overline{\xi})^2 \tag{10}$$

更方便，因为由 (9)

$$Es^{*2}=\frac{n}{n-1}\frac{n-1}{n}\sigma^2=\sigma^2. \tag{11}$$

(二) **正态母体情况**. 当 ξ 的分布一般时， $g(\xi_1,\cdots,\xi_n)$ 的精确分布很难求得. 但在正态分布情形，有些重要的结果，现在就来叙述.

定理 1 设 ξ 有正态分布 $N(a,\sigma)$, 则容量为 n 的子样 $(\xi_1,\cdots,\xi_n)$ 的平均值 $\overline{\xi}$ 及方差 s^2 相互独立，而且 $\overline{\xi}$ 有 $N\left(a,\frac{\sigma}{\sqrt{n}}\right)$ 分布， $\frac{ns^2}{\sigma^2}$ 有自由度为 $n-1$ 的 χ^2 分布.

注 1 以后记自由度为 k 的 χ^2 分布为 $\chi^2(k)$.

证 先在 $a=0$ 的假定下证明所需结论. 引进新的随机变数 $y_1,\cdots,y_n$;

$$\left.\begin{aligned}
y_1&=\frac{1}{\sqrt{1\cdot 2}}(\xi_1-\xi_2),\\
y_2&=\frac{1}{\sqrt{2\cdot 3}}(\xi_1+\xi_2-2\xi_3),\\
&\cdots\cdots\cdots,\\
y_{n-1}&=\frac{1}{\sqrt{(n-1)n}}(\xi_1+\xi_2+\cdots+\xi_{n-1}-(n-1)\xi_n),\\
y_n&=\frac{1}{\sqrt{n}}(\xi_1+\xi_2+\cdots+\xi_{n-1}+\xi_n).
\end{aligned}\right\} \tag{12}$$

我们来证明： $y_1,\cdots,y_n$ 独立，有相同分布为 $N(0,\sigma)$.

实际上，以 a_{ij} 表 (12) 中第 i 个方程内 ξ_j 的系数，易见

$$\left.\begin{aligned}
\sum_{k=1}^{n}a_{ik}a_{jk}&=\delta_{ij},\\
\sum_{k=1}^{n}a_{ki}a_{kj}&=\delta_{ij},
\end{aligned}\right\} \tag{13}$$

其中 $\delta_{ij}=1$ 或 0, 视 $i=j$ 或 $i\neq j$ 而定，因而 $(y_1,\cdots,y_n)$ 是自 $(\xi_1,\cdots,\xi_n)$ 经正交变换而来，既然正交变换保留长度不变，故

$$\sum_{i=1}^{n} y_i^2=\sum_{i=1}^{n}\xi_i^2,$$

而且 (12) 中变换行列式的值为 1.

注意 $(\xi_1,\cdots,\xi_n)$ 独立，有相同分布为 $N(0,\sigma)$, 故它们的联合分布是 n 维正态的，密度函数为

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\prod_{i=1}^{n}e^{-\frac{x_i^2}{2\sigma^2}}.$$

经过此正交变换后，$y_1,\cdots,y_n$ 的联合分布也是 n 维正态的 (见 §2.12(三)).

由 (12)(13) 易见

$$Ey_i=0,Ey_i^2=\sigma^2,Ey_iy_j=0,(i\neq j).\tag{14}$$

由不相关性立得 $y_1,\cdots,y_n$ 的独立性 (§2.12). 既然正态分布的边沿分布也是正态的，故由 (14) 还可见 y_i 有 $N(0,\sigma)$ 分布.

(12) 最后一式表示 $y_n=\overline{\xi}\sqrt{n}$, 由此得

$$ns^2=\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2=\sum_{i=1}^{n}\xi_i^2-n\overline{\xi}^2=\sum_{i=1}^{n}y_i^2-y_n^2=\sum_{i=1}^{n-1}y_i^2,\tag{15}$$

$$\frac{ns^2}{\sigma^2}=\sum_{i=1}^{n-1}\left(\frac{y_i}{\sigma}\right)^2\tag{16}$$

因为 $\frac{ns^2}{\sigma^2}$ 是 $n-1$ 个独立的具有相同分布 $N(0,1)$ 的随机变数 $\frac{y_i}{\sigma}$ 的平方的和，故由 §2.8 定理 2 知 $\frac{ns^2}{\sigma^2}$ 有 $\chi^2(n-1)$ 分布.

最后，由于 $y_1,\cdots,y_{n-1},y_n$ 独立，由 §3.3 引理 1, 知 $s^2=\frac{\sigma^2}{n}\sum\limits_{i=1}^{n-1}\left(\frac{y_i}{\sigma}\right)^2$ 与 $\overline{\xi}=\frac{y_n}{\sqrt{n}}$ 独立，于是定理在 $a=0$ 的假定下得证.

今除去此假定，考虑

$$\zeta=\xi-a,\quad \zeta_i=\xi_i-a.$$

显然 ζ 有 $N(0,\sigma)$ 分布，故可将所得结果用于 $\zeta_1,\cdots,\zeta_n$. 然而

$$\overline{\zeta}=\overline{\xi}-a,\quad \frac{1}{n}\sum_{i=1}^{n}(\zeta_i-\overline{\zeta})^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2=s^2,$$

由 $\overline{\zeta},s^2$ 的独立性立得 $\overline{\xi},s^2$ 的独立性，既然 $\overline{\zeta}$ 有 $N(0,\sigma)$ 分布，可见 $\overline{\xi}$ 有 $N(a,\sigma)$ 分布. □

注 2　一眼看来，$\overline{\xi}$ 与 s^2 的相互独立性使人惊异，因为 $s^2=\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\overline{\xi})^2$ 是 $\overline{\xi}$ 的函数. 然而这 n 个平方项 **不是** 独立的，这是由于，它们之间有关系式 $\sum\limits_{i=1}^{n}(\xi_i-\overline{\xi})=0$. 引进 y_i 后，(15) 把 ns^2 化为 $n-1$ 个 **独立** 随机变数的平方和，由此可见 $\frac{ns^2}{\sigma^2}$ 的 χ^2 分布有自由度为 $n-1$, 这不是偶然的.

系 1 设 ξ 有正态分布 $N(a,\sigma)$, 则随机变数

$$t=\sqrt{n-1}\,\frac{\overline{\xi}-a}{s} \tag{17}$$

有自由度为 $n-1$ 的学生分布.

注 3 以后记自由度为 k 的学生分布为 $t(k)$.

证 由定理 1 知下列二随机变数

$$X=\sqrt{n}\,\frac{\overline{\xi}-a}{\sigma},\quad Y=\frac{ns^2}{\sigma^2}$$

独立, 而且 X 有 $N(0,1)$ 分布, Y 有自由度为 $n-1$ 的 χ^2 分布. 于是由 §2.8 定理 3 知

$$t=\sqrt{n-1}\,\frac{X}{\sqrt{Y}}=\sqrt{n-1}\,\frac{\overline{\xi}-a}{s}$$

的分布是 $t(n-1)$ 分布. □

注意 (17) 中的 t 及其分布都与 σ 无关, 而 ξ 的分布则依赖于 a 及 σ, 这使我们有可能在未知 σ 时来检验关于 a 的假定, 见 §5.5.

考虑 Gamma 函数

$$\Gamma(p)=\int_0^\infty x^{p-1}e^{-x}dx,\ (p>0). \tag{18}$$

在数学分析中已看到

$$\Gamma(p+1)=p\Gamma(p), \tag{19}$$

特别, 当 p 等于正整数 n 时, 得

$$\Gamma(n+1)=n! \tag{20}$$

在 (18) 中, 作变换 $y=\frac{x}{c},(c>0)$, 得

$$\frac{\Gamma(p)}{c^p}=\int_0^\infty y^{p-1}e^{-cy}dy. \tag{21}$$

如果 $c=\alpha+\beta i$ 为复数, 但 $\alpha>0$, 由复变函数积分论中的柯西 (Cauchy) 定理, (21) 仍成立.

考虑函数

$$\varphi(x)=\begin{cases}0, & \text{当}x\leqslant 0,\\ \dfrac{b^p}{\Gamma(p)}x^{p-1}e^{-bx}, & \text{当}x>0,\end{cases} \tag{22}$$

其中 $b>0,p>0$ 为二常数. 因为 $\varphi(x)\geqslant 0$, 而且由 (21) 得

$$\int_{-\infty}^\infty \varphi(x)dx=\int_0^\infty \frac{b^p}{\Gamma(p)}x^{p-1}e^{-bx}dx=1.$$

故 $\varphi(x)$ 是密度函数, 以 $\varphi(x)$ 为密度函数的分布称为 Gamma 分布或 Γ- 分布, 它含二参数 $b>0,p>0$, 记此分布为 $\Gamma[b,p]$.

由 (21) 易得 Γ- 分布的特征函数为

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} e^{itx}\varphi(x)dx = \frac{b^p}{\Gamma(p)}\int_0^{\infty} x^{p-1}e^{-(b-it)x}dx \\ &= \frac{b^p}{\Gamma(p)}\cdot\frac{\Gamma(p)}{(b-it)^p} = \left(1-\frac{it}{b}\right)^{-p}. \end{aligned} \tag{23}$$

由 (23) 知：对 Γ- 分布，下面的加法定理成立：

$$\Gamma[b,p_1]*[b,p_2] = \Gamma[b,p_1+p_2] \tag{24}$$

Γ- 分布的两个重要特殊情况是：

(i) $p=1$ 时化为指数分布

$$\varphi(x) = \begin{cases} 0, & 当 x \leqslant 0, \\ be^{-bx}, & 当 x > 0. \end{cases} \tag{25}$$

(ii) $p=\frac{n}{2}, b=\frac{1}{2}$ 时，化为具自由度 n 的 χ^2 分布

$$\varphi(x) = \begin{cases} 0, & 当 x \leqslant 0, \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & 当 x > 0. \end{cases} \tag{26}$$

回到原来的问题，设 ξ,η 分别有正态分布为 $N(a_1,\sigma)$ 及 $N(a_2,\sigma)$, 它们的方差相同. 对 ξ 进行 n_1 次独立的观察而得 $\xi_1,\cdots,\xi_{n_1}$, 作

$$\overline{\xi} = \frac{1}{n_1}\sum_{i=1}^{n_1}\xi_i,\quad s_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(\xi_i-\overline{\xi})^2;$$

对 η 进行 n_2 次独立的观察而得 $\eta_1,\cdots,\eta_{n_2}$, 作

$$\overline{\eta} = \frac{1}{n_2}\sum_{i=1}^{n_2}\eta_i,\quad s_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(\eta_i-\overline{\eta})^2.$$

定理 2 设 $\xi_1,\cdots,\xi_{n_1},\eta_1,\cdots,\eta_{n_2}$ 相互独立，则随机变数

$$T = \sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\cdot\frac{\overline{\xi}-\overline{\eta}-(a_1-a_2)}{\sqrt{n_1s_1^2+n_2s_2^2}}. \tag{27}$$

有 $t(n_1+n_2-2)$ 分布

证 由定理 1, $\frac{n_1s_1^2}{\sigma^2}$, $\frac{n_2s_2^2}{\sigma^2}$ 分别有 $\chi^2(n_1-1)$ 及 $\chi^2(n_2-1)$ 分布，由 (24) 知

$$Y = \frac{n_1s_1^2+n_2s_2^2}{\sigma^2}$$

有 $\chi^2(n_1+n_2-2)$ 分布. 其次，$\overline{\xi}-\overline{\eta}-(a_1-a_2)$ 有 $N\left(0,\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$ 分布，故

$$X = \frac{\overline{\xi}-\overline{\eta}-(a_1-a_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \sqrt{\frac{n_1n_2}{n_1+n_2}}\cdot\frac{\overline{\xi}-\overline{\eta}-(a_1-a_2)}{\sigma}$$

有 $N(0,1)$ 分布，于是

$$T=\sqrt{n_1+n_2-2}\ \frac{X}{\sqrt{Y}} \tag{28}$$

有 $t(n_1+n_2-2)$ 分布. □

定理 2 可用来检查二平均值 a_1, a_2 是否相等 (见 $ 5.5).

§5.3 点 估 值

(一) 估值好坏的标准. 设随机变数 ξ 的数学期望 $a=E\xi$ 未知，试问如何根据对 ξ 的 n 次独立观察 $\xi_1,\cdots,\xi_n$ 以估计 a 的值？

由于 $a=E\xi$, 自然想到，应该用

$$\widehat{a}(\xi_1,\cdots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i \tag{1}$$

作为 a 的估值，然而，$\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 是否 a 的 **最佳** 估值？所谓"佳"的标准又是什么？

问题的一般提法如下：设 ξ 的分布函数为 $F(x,\theta_1,\cdots,\theta_k)$, 其中 $\theta_1,\cdots,\theta_k$ 是未知参数；试根据对 ξ 的 n 次独立观察 $\xi_1,\cdots,\xi_n$, 求出 θ_i 的最佳估值 $\widehat{\theta}_i(\xi_1,\cdots,\xi_n)$, 这里 $\widehat{\theta}_i(x_1,\cdots,x_n)$, $(x_j\in R_1)$ 是待求的某波雷耳可测函数.

由问题的提法可见，首先要树立估值好坏的标准. 总的想法是：希望未知参数 θ 与它的估值 $\widehat{\theta}_i(\xi_1,\cdots,\xi_n)$ 在某种意义下最为接近，具体说来有

1° 无偏性：称估值 $\widehat{\theta}=\widehat{\theta}_i(\xi_1,\cdots,\xi_n)$ 是无偏的，如果 [1]

$$E\widehat{\theta}=\theta. \tag{2}$$

例如，由 §5.2(4) 可见 $\overline{\xi}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 是数学期望 $a=E\xi$ 的无偏估值. 非无偏的估值可举方差 $\sigma^2=D\xi$ 的估值 $s^2=\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\overline{\xi})$ 为例，因为由 §5.2(9); $Es^2=\frac{n-1}{n}\sigma^2\neq\sigma^2$. 然而修改 s^2 使成 $s^{*2}=\frac{n}{n-1}s^2$ 后，则 s^{*2} 是无偏的.

无偏性远不能唯一决定估值. 例如 $\widehat{a}'=\xi_1$ 显然也是 a 的一个无偏估值.

在许多无偏的估计中，自然应以对 θ 的平均偏差较小的为好，故可引入

2° 有效性：设 $\widehat{\theta}$ 及 $\widehat{\theta}'$ 都是 θ 的无偏估值，如果

$$D\widehat{\theta}\leqslant D\widehat{\theta}', \tag{3}$$

就说 $\widehat{\theta}$ 较 $\widehat{\theta}'$ 有效；如果对固定的 n, $D\widehat{\theta}$ 的值达到极小，就称 $\widehat{\theta}$ 为 θ 的有效估值.

例如，考虑 $a=E\xi$ 的两个无偏估值 $\overline{\xi}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 及 $\widehat{a}'=\xi_1$, 由 §5.2(5) 得

$$D\overline{\xi}=\frac{\sigma^2}{n}\leqslant\sigma^2=\widehat{Da}',$$

[1] $E\widehat{\theta}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\widehat{\theta}(x_1,\cdots,x_n)\,dF(x_1,\theta)\cdots dF(x_n,\theta)$

故 $\overline{\xi}$ 较 ξ_1 有效.

甚至任取 n 个非负数 c_i, 使 $\sum\limits_{i=1}^{n} c_i = 1$, 显然

$$\widehat{a} = \sum_{i=1}^{n} c_i \xi_i$$

也是 a 的一个无偏估值，利用不等式

$$\left(\sum_{i=1}^{n} c_i\right)^2 \leqslant n \sum_{i=1}^{n} c_i^2,$$

立得

$$D\overline{\xi} = \frac{\sigma^2}{n} = \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} c_i\right)^2 \leqslant \sigma^2 \left(\sum_{i=1}^{n} c_i^2\right) = D\widehat{a},$$

可见 $\overline{\xi}$ 较一切 $\sum\limits_{i=1}^{n} c_i\xi_i$ 有效. 由于这种种原因，我们乐意用 $\frac{1}{n}\sum\limits_{i=1}^{n} \xi_i$ 作为 $E\xi$ 的估值.

注意 $\widehat{\theta} = \widehat{\theta}(\xi_1, \cdots, \xi_n)$ 仍赖于子样的容量 n, 有必要强调 n 时，宜记 $\widehat{\theta}$ 为 $\widehat{\theta}_n$, 无偏性与有效性适用于任一固定的 n.

自然希望，当子样的容量 n 越大时，对 θ 的估值越精确，因而引进下面的"一致性"定义，它适应于当 n 充分大的子样.

3° 一致性：称估值 $\widehat{\theta}_n$ 为一致的，如对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P(|\widehat{\theta}_n - \theta| > \varepsilon) = 0. \tag{4}$$

为了使 $\widehat{\theta}_n$ 是 θ 的一致估值，只需

$$\lim_{n\to\infty} E|\widehat{\theta}_n - \theta|^r = 0. \tag{5}$$

对某 $r > 0$ 成立. 实际上由马尔科夫不等式

$$P(|\widehat{\theta}_n - \theta| > \varepsilon) \leqslant \frac{E|\widehat{\theta}_n - \theta|^r}{\varepsilon^r} \to 0, \ (n \to \infty). \tag{6}$$

举几个简单的例：

由大数定理

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} \xi_i - E\xi\right| > \varepsilon\right) \to 0, \ (n \to \infty), \tag{7}$$

故 $\overline{\xi}_n = \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i$ 是 $E\xi$ 的一致估值.

其次，由于 $s_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i^2 - \overline{\xi}_n^2$, 故如 $D\xi = \sigma^2$ 有穷，在 (7) 中以 ξ_i^2 代 ξ_i 后，知 $\frac{1}{n}\sum\limits_{i=1}^{n} \xi_i^2$ 依概率收敛于 $E\xi^2$, 而 $\overline{\xi}_n^2$ 依概率收敛于 $(E\xi)^2$, 从而 $s_n^2 \to \sigma^2$, (P) [1], 这表示 s_n^2 是 σ^2 的一致估值. 因为 $\frac{n}{n-1} \to 1$, 故 $s_n^{*2} = \frac{n}{n-1} s_n^2$ 也是 σ^2 的一致估值.

[1] 如用强大数定理，甚至有 $\overline{\xi} \to E\xi$ (a.s.), $s_n^2 \to \sigma^2$ (a.s.)

类似地可以证明 $\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^r$ 是 $E\xi^r$ 的无偏一致估值， $r\geqslant 1$.

上述无偏性，有效性，一致性是对估值的基本要求，除了这些以外，还有种种标准，不去深究了.

(二) **求估值的方法**. 至今已看到：$\bar{\xi}, s^{*2}, \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^r (r\geqslant 1)$ 分别是 $E\xi, D\xi$ 及 $E\xi^r$ 的无偏、一致估值，而 s^2 是 $D\xi$ 的一致估值，因而对这些数字特征的点估值是简单的.

例如，正态分布 $N(a,\sigma)$ 中的 a(数学期望),σ^2(方差); 普阿松分布中的 λ (数学期望); 二项分布 $B(n,p)$ 中的 np(数学期望) 等等的估值都可由上求出.

一事件 A 出现的概率 $P(A)$ 可用频率 $\frac{f_n}{n}$ 来作估值，实际上，令 $\xi=\chi_A$, 其中 $\chi_A(\omega)=1$ 或 0, 视 $\omega\in A$ 或 $\omega\bar{\in}A$ 而定，则 $P(A)=E\xi$, 而 $\frac{f_n}{n}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$.

至于一般的参数 θ 应如何求它的估值呢？常用的有两种方法：矩法与最大似然法.

(A) 矩法：

矩法的本质在于以子样的 r 阶矩作为 $E\xi^r$ 的估值.

设 ξ 的分布函数的形状 $F(x,\theta_1,\cdots,\theta_k)$ 已知，但参数 $\theta_1,\cdots,\theta_k$ 未知，我们假定 ξ 的 k 阶矩 m_k 存在，因而 ξ 的 r 阶矩

$$m_r=m_r(\theta_1,\cdots,\theta_k)=\int_{-\infty}^{\infty}x^r dF(x,\theta_1,\cdots,\theta_k),\ r=1,2,\cdots,k$$

是 $\theta_1,\cdots,\theta_k$ 的函数. 考虑对 ξ 的 n 次独立观察 $\xi_1,\cdots,\xi_n$, 作此子样的 r 阶矩

$$\widehat{m}_r=\frac{1}{n}\sum_{i=1}^{n}\xi_i^r.$$

在 k 个方程

$$m_r(\theta_1,\cdots,\theta_k)=\widehat{m}_r,\ r=1,2,\cdots,k \tag{8}$$

中, 视 $\widehat{m}_r$ 为已知, 视 $\theta_1,\cdots,\theta_k$ 为未知, 一般可解得 $\widehat{\theta}_1=\widehat{\theta}_1(\xi_1,\cdots,\xi_n),\cdots,\widehat{\theta}_k=\widehat{\theta}_k(\xi_1,\cdots,\xi_n)$, 我们就以 $\widehat{\theta}_j$ 作为 θ_j 的估值， $j=1,\cdots,k$.

例 1 设 ξ 有均匀分布，密度为

$$\varphi(x)=\begin{cases}\dfrac{1}{\theta_2}, & \text{如} x\in[\theta_1,\theta_1+\theta_2],\\ 0, & \text{如} x\bar{\in}[\theta_1,\theta_1+\theta_2],\end{cases} \tag{9}$$

其中 θ_1,θ_2 为未知参数，由于

$$m_1=E\xi=\theta_1+\frac{\theta_2}{2},$$
$$m_2=E\xi^2=\theta_1^2+\theta_1\theta_2+\frac{\theta_2^2}{3},$$

自 (8) 得二方程

$$\begin{aligned}&\theta_1+\frac{\theta_2}{2}=\bar{\xi},\\ &\theta_1^2+\theta_1\theta_2+\frac{\theta_2^2}{3}=\frac{1}{n}\sum_{i=1}^{n}\xi_i^2.\end{aligned} \tag{10}$$

解 (10) 即得 θ_2,θ_1 的估值分别为

$$\widehat{\theta}_2=\sqrt{12}\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2-(\bar{\xi})^2\right\}^{1/2}=\sqrt{12}\,s_n,$$

$$\widehat{\theta}_1=\bar{\xi}-\frac{\widehat{\theta}_2}{2}.$$

而且 $\widehat{\theta}_1$ 及 $\widehat{\theta}_2$ 都是一致估值，附带指出， $D\xi=\frac{\theta_2^2}{12}$.

(B) 最大似然法：

设 ξ 的分布是连续型的，密度函数 $f(x,\theta_1,\cdots,\theta_k)$ 的形状已知，但含 k 个未知参数 $\theta_1,\cdots,\theta_k$. 以 $\xi_1,\cdots,\xi_n$ 分别代入其中的 x, 将所得 n 个函数相乘而得函数

$$L(\xi_1,\cdots,\xi_n;\theta_1,\cdots,\theta_k)=\prod_{i=1}^{n}f(\xi_i,\theta_1,\cdots,\theta_k) \tag{11}$$

称函数 L 为似然函数. 当子样 $\xi_1,\cdots,\xi_n$ 固定时， L 是 $\theta_1,\cdots,\theta_k$ 的函数. 最大似然法的本质在于：取使 L 达到最大值的 $\widehat{\theta}_1,\cdots,\widehat{\theta}_k$ 作为 $\theta_1,\cdots,\theta_k$ 的估值. 注意 $\widehat{\theta}_i$ 是 $\xi_1,\cdots,\xi_n$ 的函数.

为求使 (11) 达到最大的 $\widehat{\theta}_1,\cdots,\widehat{\theta}_k$, 考虑

$$\log L=\sum_{i=1}^{n}\log f(\xi_i,\theta_1,\cdots,\theta_k).$$

取 log 后把乘法化为加法；由于 $\log x$ 是 x 的上升函数，故 L 与 $\log L$ 在相同的点上达到最大值. 因此只要解方程组

$$\frac{\partial\log L}{\partial\theta_j}=0,\quad(1\leqslant j\leqslant k). \tag{12}$$

最大似然法的直观想法是：一试验如有若干个结果 $A,B,C,\cdots$, 如果 A 已出现，那么一般说来当时试验的条件应更有利于 A 的出现[1]，故应如此选 $\theta_1,\cdots,\theta_k$ 的值，使 $\prod\limits_{i=1}^{n}f(\xi_i,\theta_1,\cdots,\theta_k)$ 达到最大，因为这样选定的 $\theta_1,\cdots,\theta_k$ 最有利于 $\xi_1,\cdots,\xi_n$ 的出现.

注意，我们这里没有讨论 (11) 的最大值能否达到， (11) 的偏导数是否有意义等等理论问题，因为这牵涉到似然函数的性质. 对于一个具体的 L, 这问题是易于讨论的.

现在来讨论离散型的 ξ, 这时应取 (11) 中的 f 为[2]

$$f(\xi_i,\theta_1,\cdots,\theta_k)=P(\xi=\xi_i).$$

然后同样地取使 (11) 中 $L(\xi_1,\cdots,\xi_n;\theta_1,\cdots,\theta_k)$ 达到最大值的 $\widehat{\theta}_1,\cdots,\widehat{\theta}_k$, 作为 $\theta_1,\cdots,\theta_k$ 的估值.

[1]举一个简单的例：设甲箱有 99 个白球 1 个黑球，乙箱有 1 个白球 99 个黑球，今随机取出一箱，再从中随机取出一球，结果取得白球. 我们自然更多地相信这球是自甲箱内取出的，因为甲箱取得白球的概率为 $\frac{99}{100}$, 远大于自乙箱取得白球的概率 1/100.

[2]$P(\xi=\xi_i)$ 与 $\theta_1,\cdots,\theta_k$ 有关

一般地，用最大似然法所得的估值的性质比用矩法所得的更好，故通常多用最大似然法.

例 2 设 ξ 有正态分布 $N(a,\sigma)$, 其中 a,σ 均未知，对子样 $\xi_1,\cdots,\xi_n$, 似然函数为

$$L=\prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(\xi_i-a)^2/2\sigma^2}\right].$$

方程组 (12) 化为

$$\frac{\partial\log L}{\partial a}=\sum_{i=1}^{n}\frac{\xi_i-a}{\sigma^2}=0,$$

$$\frac{\partial\log L}{\partial\sigma}=\frac{1}{\sigma^3}\sum_{i=1}^{n}(\xi_i-a)^2-\frac{n}{\sigma}=0.$$

由此二方程解得

$$a=\overline{\xi},\quad \sigma^2=s^2$$

这说明由最大似然法所求得的 a 及 σ^2 的估值为 $\overline{\xi}$ 及 s^2.

例 3 设 ξ 有普阿松分布，试用最大似然法估计分布中的参数 λ. 取子样 $\xi_1,\cdots,\xi_n$, 为方便起见，令 $r=\max(\xi_1,\cdots,\xi_n)$, 又以 f_i 表等于 i 的 ξ_j 的个数，则 $f_0+f_1+\cdots+f_r=n$. 因为 $P(\xi=i)=e^{-\lambda}\cdot\frac{\lambda^i}{i!}$, 故似然函数为

$$L=\prod_{i=0}^{r}\left(e^{-\lambda}\frac{\lambda^i}{i!}\right)^{f_i}.$$

方程 (12) 化为

$$\frac{\partial\log L}{\partial\lambda}=\sum_{i=0}^{r}f_i\left(\frac{i}{\lambda}-1\right)=0.$$

由 $\sum\limits_{i=0}^{r}if_i=\sum\limits_{i=1}^{n}\xi_i$ 得 λ 的估值为

$$\sum_{i=0}^{r}if_i\Big/\sum_{i=0}^{r}f_i=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\overline{\xi}.$$

例 4 (钓鱼问题) 为了估计湖中的鱼数 N, 同时自湖中钓出 r 条，做上记号后放回池中；然后再自湖中同时钓出 s 条，结果发现这 s 条中有 ξ_1 条标有记号. 这里 N 是未知常数， r,s 是已知常数，试问应如何估计 N 的值？第二次钓出的有记号的鱼数 ξ 是随机变数， ξ 有超几何分布

$$P(\xi=\xi_1)=\frac{\binom{r}{\xi_1}\binom{N-r}{s-\xi_1}}{\binom{N}{s}},\tag{13}$$

ξ_1 为整数， $\max(0,s-N+r)\leqslant\xi_1\leqslant\min(r,s)$.

以 $L(\xi_1,N)$ 表 (13) 右方值，应取使 $L(\xi_1,N)$ 达到最大值的 $\widehat{N}$ 作为 N 的最大似然估值. 为此，用对 N 求导数的方法相当困难，我们改用下法，考虑比值

$$R(\xi_1,N)=\frac{L(\xi_1,N)}{L(\xi_1,N-1)},$$

以 (13) 右方值代入 L 中后，可以看出：当而且只当 $N < \frac{rs}{\xi_1}$ 时 $R(\xi_1, N) > 1$, 即 $L(\xi_1, N) > L(\xi_1, N-1)$; 当而且只当 $N > \frac{rs}{\xi_1}$ 时， $R(\xi_1, N) < 1$, 即 $L(\xi_1, N) < L(\xi_1, N-1)$. 故 $L(\xi_1, N)$ 在 $\frac{rs}{\xi_1}$ 附近取最大值，于是

$$\widehat{N} = \left[\frac{rs}{\xi_1}\right].$$

§5.4 区 间 估 值

(一) **置信区间**. 在上节中，我们用子样的函数 $\widehat{\theta} = \widehat{\theta}(\xi_1, \cdots, \xi_n)$ 来作未知参数 θ 的估值，当 $\xi_1, \cdots, \xi_n$ 固定时， $\widehat{\theta}$ 是一个数 (或点). 对点估值，即使它是无偏及一致的，也没有回答下列问题：某随机区间 (即端点为随机变数的区间) 包含 θ 的概率是多少？

这一节的目的是：对于未知参数 θ, 试找出两个子样函数 $\widehat{\theta}_1 = \widehat{\theta}_1(\xi_1, \cdots, \xi_n)$, $\widehat{\theta}_2 = \widehat{\theta}_2(\xi_1, \cdots, \xi_n)$, 使随机区间 $(\widehat{\theta}_1, \widehat{\theta}_2)$ 含 θ 的概率为已给值 $1-\varepsilon$:

$$P(\widehat{\theta}_1 < \theta < \widehat{\theta}_2) = 1 - \varepsilon. \tag{1}$$

称 $(\widehat{\theta}_1, \widehat{\theta}_2)$ 为 θ 的置信区间， $1-\varepsilon$ 为此区间的置信系数.

(二) **正态分布母体情形**. 设 ξ 有 正态 分布 $N(a, \sigma)$, 在这假定下容易解决求 a 及 σ 的置信区间问题.

1° 估计 $a = E\xi$

在 §5.2 中已知：随机变数

$$t = \sqrt{n-1}\,\frac{\overline{\xi} - a}{s} \tag{2}$$

有 $t(n-1)$ 分布. 对已给的正数 p (这里及以后类似的 p, 都应满足 $0 \leqslant \frac{p}{100} \leqslant 1$) , 由于此分布是连续型的，可以找到正数 t_p, 使

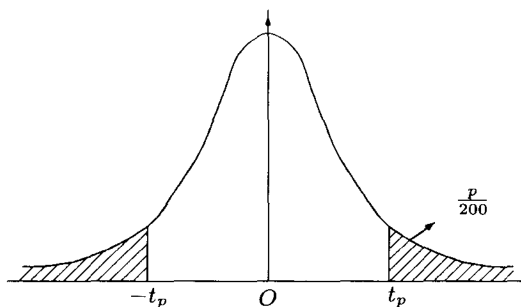


图 1 学生分布密度函数

$$P\Big(-t_p < \sqrt{n-1}\,\frac{\overline{\xi} - a}{s} < t_p\Big) = 1 - \frac{p}{100}, \tag{3}$$

此 t_p 的值可以从 t 分布的表查出，改写 (3) 为

$$P\Big(\overline{\xi} - t_p\frac{s}{\sqrt{n-1}} < a < \overline{\xi} + t_p\frac{s}{\sqrt{n-1}}\Big) = 1 - \frac{p}{100}, \tag{4}$$

即得 a 的置信区间为 $(\overline{\xi} - t_p\frac{s}{\sqrt{n-1}}, \overline{\xi} + t_p\frac{s}{\sqrt{n-1}})$, 它的置信系数为 $1 - \frac{p}{100}$. 通常取 $p = 5$, 或 1, 或 0.1, 随问题所要求的精确程度而定.

(4) 式的实际意义如下：对 m 个子样

$$\xi^{(i)} = (\xi_1^{(i)}, \cdots, \xi_n^{(i)}),\ i = 1, \cdots, m,$$

可得 m 个区间

$$I^{(i)} = \Big(\overline{\xi^{(i)}} - t_p\frac{s(\xi_1^{(i)}, \cdots, \xi_n^{(i)})}{\sqrt{n-1}},\ \overline{\xi^{(i)}} + t_p\frac{s(\xi_1^{(i)}, \cdots, \xi_n^{(i)})}{\sqrt{n-1}},\Big)$$

$(i=1,\cdots,m)$. 由强大数定理，当 m 充分大后，大约有 $(1-\frac{p}{100})m$ 个这种区间包含 a; 例如当 $p=1, m=1000$ 时，大约有 990 个包含 a, 不包含 a 的只有 10 个.

注意，对于已给的 p, 置信区间不是唯一的，例如，以 $t_{n-1}(x)$ 表 $t(n-1)$ 分布的密度函数，只要选二数 $\alpha<\beta$, 使满足

$$P\Big(\alpha<\sqrt{n-1}\,\frac{\bar{\xi}-a}{s}<\beta\Big)=\int_\alpha^\beta t_{n-1}(x)\,dx=1-\frac{p}{100}, \tag{5}$$

那么 $(\bar{\xi}-\beta\frac{s}{\sqrt{n-1}},\bar{\xi}-\alpha\frac{s}{\sqrt{n-1}})$ 就是一置信区间. 在许多具有同样置信系数的置信区间中，我们自然愿意挑选长度最短的.

对固定的 n, 当 p 变小时， t_p 变大，故置信区间的长度也越大.

例 1 在稳定生产的情况下，某工厂出品的电灯泡的使用时数 ξ 一般可假定有正态分布 $N(a,\sigma)$, 其中 a,σ 未知，观察 $n=20$ 个灯泡后得 20 个使用时数 $\xi_1,\cdots,\xi_{20}$, 由此子样算得 $\bar{\xi}=1832$, $s=497$, 试求电灯泡的平均使用时数 a 的置信区间 [1].

解 这里自由度为 $n-1=19$, 查 $t(19)$ 分布表后得

p%	t_p	$(\bar{\xi}-t_p\frac{s}{\sqrt{n-1}},\bar{\xi}+t_p\frac{s}{\sqrt{n-1}})$
5%	2.093	(1953, 2071)
1%	2.861	(1506, 2158)
0.1%	3.883	(1389, 2275)

注：如果 σ 已知，可以用较简单的方法来估计 a, 实际上，考虑随机变数 $\sqrt{n}\,\frac{\bar{\xi}-a}{\sigma}$, 由于它是 $\xi_1,\cdots,\xi_n$ 的线性组合，故由 §2.11(44) 知它有正态分布；根据 §5.2 例 1, 还知它有 $N(0,1)$ 分布. 因此，对已给正数 p, $0<\frac{p}{100}<1$, 由正态分布表，存在 $\lambda_p>0$, 使

$$P\big(-\lambda_p<\sqrt{n}\,\frac{\bar{\xi}-a}{\sigma}<\lambda_p\big)=1-\frac{p}{100},$$

亦即

$$P\big(\bar{\xi}-\lambda_p\frac{\sigma}{\sqrt{n}}<a<\bar{\xi}+\lambda_p\frac{\sigma}{\sqrt{n}}\big)=1-\frac{p}{100}.$$

2° 估计. $\sigma^2=D\xi$.

方法与 1° 完全一样，回顾上段可见估计 $a=E\xi$ 方法的本质在于能找到一个含未知参数 a 及子样的随机变数 $t=\sqrt{n-1}\,\frac{\bar{\xi}-a}{s}$, 使 t 的分布完全确定，而且不依赖于 a. 为了估计 σ^2, 我们考虑随机变数 $\frac{ns^2}{\sigma^2}$ 以代替 $\sqrt{n-1}\,\frac{\bar{\xi}-a}{s}$, 注意，由 §5.2 知："$\frac{ns^2}{\sigma^2}$ 有 $\chi^2(n-1)$ 分布". 对已给的 $\frac{p}{100}$, 利用 χ^2 分布表，可找到二数 $\chi^2_{p'}$ 及 $\chi^2_{p''}$, 使

$$P\Big(\chi^2_{p'}<\frac{ns^2}{\sigma^2}<\chi^2_{p''}\Big)=1-\frac{p}{100}, \tag{6}$$

满足这条件的 $\chi^2_{p'},\chi^2_{p''}$ 有无穷多个，一般用下二式选取 $\chi^2_{p'}$ 及 $\chi^2_{p''}$, 以使置信区间的长度较短：

$$P\Big(\frac{ns^2}{\sigma^2}>\chi^2_{p'}\Big)=1-\frac{p}{200},$$

[1]此例及以下一些例中的数据录自 [50]

$$P\left(\frac{ns^2}{\sigma^2} \geqslant \chi^2_{p''}\right) = \frac{p}{200}.$$

注意这样选取的 $\chi^2_{p'}, \chi^2_{p''}$ 唯一.

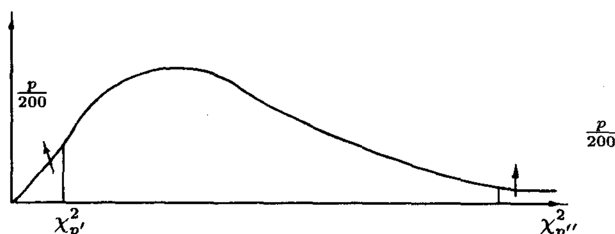


图 2 自由度大于 2 的 χ^2 分布密度函数图

由 (6) 立得

$$P\left(\frac{ns^2}{\chi^2_{p''}} < \sigma^2 < \frac{ns^2}{\chi^2_{p'}}\right) = 1 - \frac{p}{100}, \tag{7}$$

$$P\left(\frac{\sqrt{n}s}{\chi_{p''}} < \sigma < \frac{\sqrt{n}s}{\chi_{p'}}\right) = 1 - \frac{p}{100}. \tag{8}$$

3° 估计二平均值的差.

设 ξ, η 为二随机变数，各有正态分布为 $N(a_1, \sigma), N(a_2, \sigma)$, 方差相等. 试根据对 ξ 的 n_1 次独立观察 $(\xi_1, \cdots, \xi_{n_1})$, 及对 η 的 n_2 次独立观察 $(\eta_1, \cdots, \eta_{n_2})$ 以估计 $a_1 - a_2$, 我们假定 $(\xi_1, \cdots, \xi_{n_1}, \eta_1, \cdots, \eta_{n_2})$ 也独立. 令

$$s_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(\xi_i - \overline{\xi})^2, \quad s_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(\eta_i - \overline{\eta})^2$$

$$\overline{\xi} = \frac{1}{n_1}\sum_{i=1}^{n_1}\xi_i, \quad \overline{\eta} = \frac{1}{n_2}\sum_{i=1}^{n_2}\eta_i,$$

由 §5.2 知

$$T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{(\overline{\xi} - \overline{\eta}) - (a_1 - a_2)}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \tag{9}$$

有 $t(n_1 + n_2 - 2)$ 分布. 于是用同样方法，对已给 $p > 0$, 取 t_p 使

$$P(|T| < t_p) = 1 - \frac{p}{100} \tag{10}$$

后，以 (9) 中 T 代入上式，经改写后即得

$$P\left(\overline{\xi} - \overline{\eta} - t_p\sqrt{\frac{(n_1+n_2)(n_1 s_1^2 + n_2 s^2)}{n_1 n_2 (n_1 + n_2 - 2)}} < a_1 - a_2 \right.$$
$$\left. < \overline{\xi} - \overline{\eta} + t_p\sqrt{\frac{(n_1+n_2)(n_1 s_1^2 + n_2 s^2)}{n_1 n_2 (n_1 + n_2 - 2)}}\right) = 1 - \frac{p}{100}. \tag{11}$$

(三) **一般母体情形**. 上段中的结果不论对大子样或小子样都适用，原因在于 (2),(9) 中的 t,T 及 $\frac{ns^2}{\sigma^2}$ 的精确分布已知，而这又来自 ξ 有 **正态** 分布的假定. 如果除去这一假定，通常精确分布很难找到而只好乞援于当 $n\to\infty$ 时的极限分布，这样就需要相当大的子样. 故这一段中的结果主要适用于大子样，我们只限于讨论 $a=E\xi$ 的区间估计问题以例示一般.

设随机变数 ξ 有非零而有穷的方差 $D\xi$, 由 §5.2 例 1 知 $\bar{\xi}$ 为 $(a,\frac{\sigma}{\sqrt{n}})$ 渐近正态的，即

$$\lim_{n\to\infty}P\left(\frac{\bar{\xi}_n-a}{\frac{\sigma}{\sqrt{n}}}\leqslant x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}\,dz, \tag{12}$$

其中 $\bar{\xi}_n=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 为子样 $(\xi_1,\cdots,\xi_n)$ 的平均值. 因而当 n 充分大时，可认为随机变数

$$u_n=\sqrt{n}\,\frac{\bar{\xi}_n-a}{\sigma} \tag{13}$$

的分布 **接近** 于 (但未必重合于)$N(0,1)$ 分布，对已给的数 p, 由正态分布表可找到正数 λ_p, 使满足

$$P\Big(-\lambda_p<\sqrt{n}\,\frac{\bar{\xi}_n-a}{\sigma}<\lambda_p\Big)\approx\frac{1}{\sqrt{2\pi}}\int_{-\lambda_p}^{\lambda_p}e^{-\frac{z^2}{2}}\,dz=1-\frac{p}{100}, \tag{14}$$

故

$$P\Big(\bar{\xi}_n-\lambda_p\,\frac{\sigma}{\sqrt{n}}<a<\bar{\xi}_n+\lambda_p\frac{\sigma}{\sqrt{n}}\Big)\approx 1-\frac{p}{100}. \tag{15}$$

上式表示，当 σ 已知而且子样的容量 n 充分大时，可近似地取

$$\Big(\bar{\xi}_n-\lambda_p\frac{\sigma}{\sqrt{n}},\bar{\xi}_n+\lambda_p\frac{\sigma}{\sqrt{n}}\Big) \tag{16}$$

为 a 的置信区间，它的置信系数约为 $1-\frac{p}{100}$.

如果 σ 未知，由于 $s_n^2=\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\bar{\xi})^2$ 是 σ^2 的一致性估值，故通常仍用 (16) 做 a 的置信区间，不过把其中的 σ 换为 s_n, 这样一来，置信系数又因此而再产生一次误差，只能说大致为 $1-\frac{p}{100}$ 罢了.

上述结果可一般化. 分析上面的论证，可见关键在于 "$\bar{\xi}$ 为 $(a,\frac{\sigma}{\sqrt{n}})$ 渐近正态的", 因此，如 ξ 的某数字特征 b 需要估计，我们作子样的对应的数字特征 $\beta_n=\beta(\xi_1,\cdots,\xi_n)$, 如果 β_n 是 $(b,c(\beta_n))$ 渐近正态的，其中 $c(\beta_n)$ 是某与 b 无关的值 (可依赖于子样或其它未知参数 $d,e,\cdots$), 那么我们就可用同样的方法来近似地求 b 的置信区间，这时如 d 等已知，则只有一次误差 (由于以正态代替了 **渐近** 正态); 如 d 等未知，它们是 ξ 的某些数字特征 (如 ξ 的方差 $\cdots$), 那么就可以用子样的同样的数字特征 (如子样的方差 $\cdots$) 来代替 d 等，因而出现了两次误差.

§5.5 假 设 检 验

(一) 基本思想. 现在来研究 §5.1(一) 中提出的第三个问题：根据对随机变数 ξ 的 n 次独立观察 $\xi_1,\cdots,\xi_n$, 试用统计方法，来检验关于 ξ 的某假设 H 是否正确. 我们主要只限于讨论如下的几种假设：

(i) 已知 ξ 有正态分布 $N(a,\sigma)$, 其中 a,σ 未知，要检验假设 H: "a 等于某常数 a_0" 或 "σ 等于某常数 σ_0", 这时 H 可简写为 "$a=a_0$" 或 "$\sigma=\sigma_0$".

(ii) ξ 的分布函数 $F(x)$ 未知，要检验 H: "$F(x)$ 重合于某分布函数 $F_0(x)$", 亦即 H 为 "$F(x)=F_0(x)$".

假设检验基本思想的根据是："小概率事件 (即概率很小的事件) 在一次试验中几乎是不可能出现的". 详细解释一下：设有某 H 需要检验，我们先假定 H 是正确的，在此假定下，某事件 A 的概率很小，例如 $P(A)=\frac{1}{100}$. 经过一次试验后，如果 A 出现，那么便出现了一个小概率事件 A. 但如上所说，小概率事件在一次试验中几乎是不可能出现的，而现在居然出现了，这就不能不使人怀疑 H 的正确性，因而自然要否定 H. 反之，如果 A 不出现，一般就肯定 H, 或保留 H, 留待经过几次试验后再作结论.

举一个例：箱中有白球或黑球，总数为 100, 假设 H: "其中 99 个是白的", 现在用上面的想法来检验 H 的正确性. 暂设 H 正确，那么从箱中任取一球，"得黑球" 的概率为 $\frac{1}{100}$, 故是一小概率事件. 今如居然抽得一黑球，那么自然使人要否定 H, 就是说，白球的个数不是 99.

采用上述方法时，可能犯两类错误：第一，H 本来是正确的，但我们却错误地否定了它，这种 "弃真" 的错误称为第一类错误. 由于我们只是在 A 出现时才否定 H, 故犯第一类错误的概率 α 等于 $P(A|H)$, 即在 H 成立的条件下 A 的条件概率. 其次，H 本来不正确，但我们却肯定了它，于是犯了 "取伪" 的错误，称为第二类错误，犯第二类错误的概率记为 β. 在上例中，$\alpha=\frac{1}{100}$.

如果只有两种假设 H 及 H' 是可能的，就是说，H,H' 中必有一正确，但不能肯定到底是哪一个，这时易见犯第二类错误 (即本来正确 H', 但错误地接受 H) 的概率 $\beta=P(\overline{A}|H')$.

自然希望，α,β 越小越好.

注意，上面的检验法完全取决于小概率事件 A 的选择，由于 A 的选法一般是多种的，故检验法也是多种的. 根据上面所说的，应该取使得 α,β 都最小的 A. 通常，要 α,β 都最小是困难的，于是固定其一，例如固定 $\alpha=\alpha_0,\alpha_0$ 是给定的常数，然后在一切满足 $P(A|H)\leqslant\alpha_0$ 的 A 中，选取一个使 β 取极小的 A_0. A_0 所决定的检验法，比其它满足 $P(A|H)\leqslant\alpha_0$ 的 A 所决定的要好，这样便产生了检验法好坏的标准.

称给定的 α_0 为信度. 以后我们否定 H, **总是相对于已给信度而言的**, 这时弃真的概率即 α_0. 信度到底应该给得多大？应随具体问题的要求而定，如果我们宁愿弃真的概率小些，那么便应取较小的 α_0.

由上面的讨论可见，构造一个检验法的关键在于选定一小概率事件 A, 然而 A 的选法又是怎样的呢？一般选法如下：选定一个 n 元波雷耳可测函数 $g(x_1,\cdots,x_n)$, 对已给信度 α_0, 取一个一维波雷耳可测集 B, 使在 H 正确的条件下，

$$P(g(\xi_1,\cdots,\xi_n)\in B)\leqslant\alpha_0.$$

因而事件 $A=(g(\xi_1,\cdots,\xi_n)\in B)$ 的概率不大于 α_0, 当 α_0 甚小时，A 就是小概率事件. 这样一来，问题就主要归结为函数 g 的选择. 函数 g 至少要具备的条件是：$g(\xi_1,\cdots,\xi_n)$ 的分布 (或当 $n\to\infty$ 时的极限分布) 已知，这样才有可能算出 (或近似地算出) 上式右方的概率 $P(g(\xi_1,\cdots,\xi_n)\in B)$. 函数 g 的选择是比较困难的问题，以后会看到一些如何选取的例子.

称随机变数 $g(\xi_1,\cdots,\xi_n)$ 为统计量.

(二) 对参数的假设检验. 假设 ξ 有正态分布 $N(a,\sigma)$, 现在来检验关于未知参数 a,σ 的一些假定 H.

1° H: "$a=a_0$".

设 H 正确而考虑统计量

$$t=\sqrt{n-1}\,\frac{\bar{\xi}-a_0}{s}. \tag{1}$$

由 §5.4(二)1°, 知 t 有 $t(n-1)$ 分布，故对 $p>0$, 由 §5.4(3)

$$P\Big(-t_p<\sqrt{n-1}\,\frac{\bar{\xi}-a_0}{s}<t_p\Big)=1-\frac{p}{100}. \tag{2}$$

因而当 p 相当小时，例如 $p=5$ 或 $p=1$, 则

$$A=\Big(\Big|\sqrt{n-1}\,\frac{\bar{\xi}-a_0}{s}\Big|\geqslant t_p\Big) \tag{3}$$

是一小概率事件，概率为 $\frac{p}{100}$. 以子样值 $\xi_1,\cdots,\xi_n$ 代入

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\quad s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2},$$

如果 (3) 中右方不等式成立，则出现了小概率事件 A 而否定 H; 否则可接受 [1] H.

例 1 继续考虑 §5.4 例 1, 试检验假定 H: "电灯泡的平均使用时数为 $a=2000$ 小时". 在那里已设 $\bar{\xi}=1832$, $s=497$, $n=20$, 代入 (1) 后得 $t=-1.473$, 由于 $|t|<2.093=t_5$ (参看那里的表),(3) 中小概率事件不出现，故相对于信度 $\frac{5}{100}$ 而言，应接受 H. 注意信度 $\frac{p}{100}$ 越小，则 t_p 越大，例如. $t_1=2.861$, $t_{0.1}=3.383$.

2° H : "$\sigma=\sigma_0$".

理论与 1° 全同，只要利用 §5.4(二)2°, 考虑统计量 $\frac{ns^2}{\sigma^2}$, 由 §5.4(6) 得

$$P\Big(\chi^2_{p'}<\frac{ns^2}{\sigma^2}<\chi^2_{p''}\Big)=1-\frac{p}{100}, \tag{4}$$

故事件 A

$$A=\Big(\frac{ns^2}{\sigma^2}\leqslant\chi^2_{p'}\Big)\cup\Big(\frac{ns^2}{\sigma^2}\geqslant\chi^2_{p''}\Big) \tag{5}$$

有概率 $\frac{p}{100}$.

---

[1] 或暂不作结论，留待进一步检验，附带指出：如 σ 已知，要检验 H: $a=a_0$, 可以采用较简单的统计量 $\sqrt{n}\,\frac{\bar{\xi}-a_0}{\sigma}$, 参看 §5.4 例 1 下的注.

3° 设 ξ 有正态分布 $N(a_1,\sigma)$, η 有正态分布 $N(a_2,\sigma)$, 试检验假定 H: $a_1=a_2$.

注意，我们已假定了 ξ_1,ξ_2 有相同的方差.

为此，利用 §5.4(二)3°, 以 §5.4(9) 中的 T 作为统计量，如 H 正确，则 T 化为

$$T=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\cdot\frac{\overline{\xi}-\overline{\eta}}{\sqrt{n_1s_1^2+n_2s_2^2}},\tag{6}$$

而且 T 应有 $t(n_1+n_2-2)$ 分布. 对 $p>0$, 可得 t_p 使

$$P(|T|<t_p)=1-\frac{p}{100}.\tag{7}$$

因而，以子样 $(\xi_1,\cdots,\xi_{n_1})(\eta_1,\cdots,\eta_{n_2})$ 的值代入 T 后，如 $|T|\geqslant t_p$, 则相对于信度 $\frac{p}{100}$ 而言，应否定 H; 如 $|T|<t_p$, 则应接受 H.

例 2 从一批电灯泡中取出 20 个，它们的平均使用时数为 $\overline{\xi}=1832$, $s_1=497$. 再从另一批电灯泡中取出 30 个，得平均使用时间为 $\overline{\eta}=1261, s_2=501$. 设二批电灯泡的母体方差相等，试检验假定 H: "二母体的平均使用时数 a_1,a_2 相等". 信度为 $\frac{0.1}{100}$.

解 这里 $n_1=20$, $n_2=30$, 由 (6) 算得 $T=3.881$. 今已给 $p=\frac{0.1}{100}$, 从自由度为 $20+30-2=48$ 的分布表查得 $t_{0.1}=3.515$, 既然 $3.881>3.515$, 故相对于信度 $\frac{0.1}{100}$ 而言，应否定 H.

例 3 设有甲、乙两种安眠药，今欲比较它们的治疗效果，以 ξ 表失眠病者服甲药后睡眠时间延长时数，以 η 表服乙药后的延长时数. 今独立观察 20 个患者，其中 10 人服甲药后的延长睡眠时数为 $\xi_1,\cdots,\xi_{10}$; 服乙药的 10 人的相应时数为 $\eta_1,\cdots,\eta_{10}$, 得表如下. 医学上可设 ξ,η 分别有 $N(a_1,\sigma),N(a_2,\sigma)$ 分布，于是问题化为检验 H: "$a_1=a_2$", 信度设为 $\frac{5}{100}$.

ξ_1	1.9	ξ_6	4.4	η_1	0.7	η_6	3.4
ξ_2	0.8	ξ_7	5.5	η_2	−1.6	η_7	3.7
ξ_3	1.1	ξ_8	1.6	η_3	−0.2	η_8	0.8
ξ_4	0.1	ξ_9	4.6	η_4	−1.2	η_9	0.0
ξ_5	−0.1	ξ_{10}	3.4	η_5	−0.1	η_{10}	2.0

解 由表算得 $\overline{\xi}=2.33,\overline{\eta}=0.75$,

$$10s_1^2=\sum_{i=1}^{10}(\xi_i-\overline{\xi})^2=36.1,\ 10s_2^2=\sum_{i=1}^{10}(\eta_i-\overline{\eta})^2=28.9,$$

$$T=\sqrt{\frac{100\cdot18}{20}}\cdot\frac{2.33-0.75}{\sqrt{36.1+28.9}}\approx1.86.$$

由 $t(18)$ 分布表查得 $t_5=2.101>1.86$, 故可接受 H.

4° 设 ξ,η 各有正态分布 $N(a_1,\sigma_1),N(a_2,\sigma_2)$, 试根据观察值 $\xi_1,\cdots,\xi_{n_1};\eta_1,\cdots,\eta_{n_2}$ 以检验假设 H: $\sigma_1=\sigma_2$. 我们设这 n_1+n_2 个随机变数是独立的.

考虑

$$s_1^{*2}=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(\xi_i-\overline{\xi})^2,\quad s_2^{*2}=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(\eta_i-\overline{\eta})^2,$$

注意 $\frac{n_1-1}{\sigma_1^2}s_1^{*2}=\frac{n_1s_1^2}{\sigma_1^2}$, 故由 §5.2 定理 1, $\frac{n_1-1}{\sigma_1^2}s_1^{*2}$ 有 $\chi^2(n_1-1)$ 分布；同理 $\frac{n_2-1}{\sigma_2^2}s_2^{*2}$ 有 $\chi^2(n_2-1)$ 分布.

今设 H 正确，以 σ^2 表 σ_1^2 与 σ_2^2 的公共值，则由 §2.8 统计量

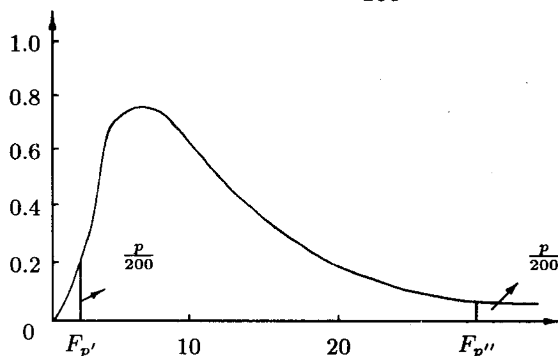
$$F=\frac{s_1^{*2}}{s_2^{*2}}=\frac{\frac{(n_1-1)s_1^{*2}}{\sigma^2}\Big/(n_1-1)}{\frac{(n_2-1)s_2^{*2}}{\sigma^2}\Big/(n_2-1)} \tag{8}$$

有参数为 n_1-1 及 n_2-1 的 F 分布的随机变数，对已给的 $p>0$, 由 F 分布表可找到二数 $F_{p'}$ 及 $F_{p''}$, 使

$$P(F>F_{p'})=1-\frac{p}{200},\quad P(F\geqslant F_{p''})=\frac{p}{200};$$

而

$$P(F_{p'}<F<F_{p''})=1-\frac{p}{100}. \tag{9}$$



F 分布密度图

以上检验法都是在 “ξ 有正态分布” 的前提下得到的. 如除去此前提，可以用 §5.4(13) 中的 $\mu_n=\sqrt{n}\,\frac{\overline{\xi}_n-a}{\sigma}$ 作为统计量，并用 §5.4(15) 来检验 H: ,“$E\xi=a$”,a 为某常数. 但此时 n 应充分大才比较有效.

(三) χ^2 **检验**. 现在来检验假设 H: “ξ 的分布函数 $F(x)$ 为 $F_0(x)$”, $F_0(x)$ 为某已知分布函数. 这里不必对 $F(x)$ 作假定，例如，不必假定已知 $F(x)$ 的形状而只要估计或检验几个参数. 因此，下面介绍的方法属于所谓非参数方法.

对 ξ 进行 N 次独立的观察而得 $\xi_1,\cdots,\xi_N$. 将 $R_1=(-\infty,\infty)$ 分为 m 个子区间 $(x_{i-1},x_i]$, 其中 $-\infty=x_0<x_1<\cdots<x_m=+\infty$(当然，$(x_{m-1},\infty]$ 理解为 (x_{m-1},∞)), 以 v_i 表 $\xi_1,\cdots,\xi_N$ 中落于 $(x_{i-1},x_i]$ 中的个数，即满足 $\xi_k\in(x_{i-1},x_i]$ 的 $k(\leqslant N)$ 的个数，显然，$\sum\limits_{i=1}^{m}v_i=N$.

今设 H: “$F(x)\equiv F_0(x)$” 正确，令 $p_i=F_0(x_i)-F_0(x_{i-1})>0$. 考虑统计量

$$\eta=\sum_{j=1}^{m}\frac{(v_j-Np_j)^2}{Np_j}=\sum_{j=1}^{m}\frac{v_j^2}{Np_j}-N \tag{10}$$

η 依赖于 N 及 m, 以下总固定 m, 注意如 H 正确，则由强大数定理 $\frac{v_i}{N}\to p_i$, (a.s.), 故当 N 甚大时，$v_i\approx Np_i$.

定理 1 (皮尔逊 [Pearson]) 如 H 正确，则

$$\lim_{N\to\infty}P\left(\sum_{j=1}^{m}\frac{(v_j-Np_j)^2}{Np_j}\leqslant x\right)=\int_0^x K_{m-1}(y)\,dy,\ (x>0) \tag{11}$$

其中

$$K_{m-1}(y)=\frac{1}{2^{\frac{m-1}{2}}\Gamma\left(\frac{m-1}{2}\right)}\,y^{\frac{m-3}{2}}e^{-\frac{y}{2}},\ (y>0) \tag{12}$$

是 $\chi^2(m-1)$ 分布的密度函数.

先对皮尔逊定理作一些解释：由中心极限定理知

$$u_j=\frac{v_j-Np_j}{\sqrt{Np_j}},\ (j=1,\cdots,m) \tag{13}$$

当 $N\to\infty$ 时有平均值为 0 的渐近 $N(0,1)$ 正态分布，由 §2.8 定理 2 容易想象 $\eta=\sum\limits_{j=1}^{m}u_j^2$ 有渐近 χ^2 分布. 注意 $u_1,\cdots,u_m$ 不是独立的，因为它们之间有关系式

$$\sum_{j=1}^{m}u_j\sqrt{p_j}=0.$$

皮尔逊定理断定， η 有 $\chi^2(m-1)$ 分布. 我们回忆 §5.2 定理 1 中的

$$\frac{ns^2}{\sigma^2}=\sum_{j=1}^{n}\left(\frac{\xi_i-\bar{\xi}}{\sigma}\right)^2$$

也有 $\chi^2(n-1)$ 分布，情况与这里的类似，参看 §5.2 注 2.

证 如 ξ 的分布函数为 $F_0(x)$, 则 $(v_1,\cdots,v_m)$ 有 m 项分布为

$$P(v_1=k_1,\cdots,v_m=k_m)=\frac{N!}{k_1!k_2!\cdots k_m!}\,p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m},$$

其中 k_i 为非负整数， $\sum\limits_{i=1}^{m}k_i=N$(参看 §2.6 例 4); 特征函数为

$$\sum_{k_1,\cdots,k_m}e^{i(t_1k_1+\cdots+t_mk_m)}\frac{N!}{k_1!\cdots k_m!}\,p_1^{k_1}\cdots p_m^{k_m}=(p_1e^{it_1}+\cdots+p_me^{it_m})^N,$$

其中 Σ 对一切满足 $k_1+\cdots+k_m=N$ 的非负整数列 $(k_1,\cdots,k_m)$ 进行. 由 §2.12 知 (13) 中的 $(u_1,\cdots,u_m)$ 的特征函数为

$$\varphi_N(t_1,\cdots,t_m)=e^{-i\sqrt{N}\sum\limits_{j=1}^{m}t_j\sqrt{p_j}}\left(p_1e^{\frac{it_1}{\sqrt{Np_1}}}+\cdots+p_me^{\frac{it_m}{\sqrt{Np_m}}}\right)^N. \tag{14}$$

利用泰勒 (Taylor) 展式

$$e^{\frac{it_j}{\sqrt{Np_j}}}-1=i\frac{t_j}{\sqrt{Np_j}}-\frac{1}{2}\frac{t_j^2}{Np_j}+o\left(\frac{1}{N}\right),$$

得

$$\log \varphi_N(t_1,\cdots,t_m)=\log e^{-i\sqrt{N}\sum\limits_{j=1}^{m}t_j\sqrt{p_j}}+N\log\left[1+\sum_{j=1}^{m}p_j\left(e^{i\frac{t_j}{\sqrt{Np_j}}}-1\right)\right]$$

$$=-i\sqrt{N}\sum_{j=1}^{m}t_j\sqrt{p_j}+N\left[\sum_{j=1}^{m}p_j\left(e^{i\frac{t_j}{\sqrt{Np_j}}}-1\right)\right]-\frac{N}{2}\left[\sum_{j=1}^{m}p_j\left(e^{i\frac{t_j}{\sqrt{Np_j}}}-1\right)\right]^2+o(1)$$

$$=-i\sqrt{N}\sum_{j=1}^{m}t_j\sqrt{p_j}+N\sum_{j=1}^{m}p_j\left(i\frac{t_j}{\sqrt{Np_j}}-\frac{t_j^2}{2Np_j}\right)-\frac{N}{2}\left[\sum_{j=1}^{m}p_j\frac{it_j}{\sqrt{Np_j}}\right]^2+o(1)$$

$$=-\frac{1}{2}\sum_{j=1}^{m}t_j^2+\frac{1}{2}\left(\sum_{j=1}^{m}t_j\sqrt{p_j}\right)^2+o(1).$$

从而

$$\lim_{N\to\infty}\varphi_N(t_1,\cdots,t_m)=e^{-\frac{1}{2}\left[\sum\limits_{j=1}^{m}t_j-\left(\sum\limits_{j=1}^{m}t_j\sqrt{p_j}\right)^2\right]} \tag{15}$$

作正交变换

$$\left.\begin{aligned}s_l&=\sum_{j=1}^{m}c_{lj}t_j\ (l=1,\cdots,m-1),\\ s_m&=\sum_{j=1}^{m}t_j\sqrt{p_j}.\end{aligned}\right\} \tag{16}$$

由正交性 $\sum\limits_{j=1}^{m}t_j^2=\sum\limits_{j=1}^{m}s_j^2$, 代入 (15) 得

$$\lim_{N\to\infty}\varphi_N(s_1,\cdots,s_m)=e^{-\frac{1}{2}\left[\sum\limits_{j=1}^{m}s_j^2-s_m^2\right]}=e^{-\frac{1}{2}\sum\limits_{j=1}^{m-1}s_j^2}.$$

右方是 $m-1$ 维正态分布的特征函数，而且有此分布的 $m-1$ 个随机变数互不相关故相互独立. 由特征函数的逆极限定理，当 $N\to\infty$ 时，$(u_1,\cdots,u_m)$ 的极限分布是此 $m-1$ 维正态分布 (逆极限定理对多维分布也成立), 从而推知当 $N\to\infty$ 时，$\eta=\sum\limits_{j=1}^{m}u_j^2$ 的极限分布是 $\chi^2(m-1)$ 分布. □

由定理 1, 当 N 充分大时，可认为 $\eta=\sum\limits_{i=1}^{m}\frac{(v_i-Np_i)^2}{Np_i}$ 有 $\chi^2(m-1)$ 分布，对已给 $p>0$, 由 χ^2 分布表可求得常数 η_p, 使

$$P(\eta>\eta_p)=\frac{p}{100}.$$

以子样值代入 η 后，如 $\eta>\eta_p$, 则相对信度 $\frac{p}{100}$ 而言，应否定 H. 在实际中用定理 1 时，通常取 m 于 7 到 14 之间.

例 4 箱中盛有 10 种球，自其中还原地随机取出 200 个后，第 i 种球共取得 v_i 个，得表如下，试检验 H: "箱内各种球的个数相同", 已给信度为 0.5%.

种　　别	v_i	Np_i	$v_i - Np_i$	$\frac{(v_i-Np_i)^2}{Np_i}$
1	35	20	15	11.25
2	16	20	−4	0.80
3	15	20	−5	1.25
4	17	20	−3	0.45
5	17	20	−3	0.45
6	19	20	−1	0.05
7	11	20	−9	4.05
8	16	20	−4	0.80
9	30	20	10	5.00
10	24	20	4	0.80
Σ	200	200	0	24.90

解　这里 $N = 200, m = 10, x_0 = -\infty, x_i = i, (i = 1, \cdots, 10)\ x_{11} = \infty$. 如 H 正确，则每次抽得第 i 种球的概率

$$p_i = \frac{1}{10}.$$

这时由附表知

$$\eta = 24.9.$$

其次，由 $\chi^2(9)$ 分布表得 $\eta_{0.005} = 23.6$, 既然 $24.9 > 23.6$, 故应否定 H. 从表上看，也可见第 1 种与第 9 种球取得过多. □

例 5　观察 $N = 2880$ 个婴儿出生时刻，得表：

小时	0	1	2	3	4	5	6	7	8	9	10	11	12
人数	127	139	143	138	134	115	127	113	126	122	121	119	130

小时	13	14	15	16	17	18	19	20	21	22	23	总计
人数	125	112	97	115	94	99	97	100	119	127	139	2880

这例说明，在 0 时内出生共 127 人，在 1 时内 139 人等等. 一眼看来，出生人数更多地集中在夜间，试问这种倾向是否显著. 精确一些，试问假定 H: "出生时刻是在 0 到 24 时内均匀分布" 是否正确？取 $p = 5$ 及 $p = 1$.

解　设 H 正确，则婴儿在每小时内出生的概率都是 $p_i = \frac{1}{24}$, 这里 $N = 2880$, 故 $Np_i = \frac{2880}{24} = 120$, 由 (10) 算得 $\eta = 40.467, m - 1 = 23$. 查 $\chi^2(23)$ 分布表得

$$\eta_5 = 35.2 < 40.467 < 41.6 = \eta_1,$$

故相对于信度 $\frac{5}{100}$ 而言，应否定 H; 对信度 $\frac{1}{100}$ 而言，应肯定 H . □

(**四**) K **检验**. 除 χ^2 检验法外，再介绍一种重要的检验法，设随机变数 ξ 的分布函数为 $F(x)$, 而且假定 $F(x)$ 是 x 的 **连续** 函数，对 ξ 进行 n 次独立观察得 $\xi_1, \cdots, \xi_n$, 根据此子样作经验分布函数 $F_n(x)$. 在 §5.1 中，格里文科定理断定

$$P\Big(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\Big) = 1.$$

这定理表示随机变数 $D_n = \sup\limits_{-\infty<x<\infty} |F_n(x) - F(x)|$ 以概率 1 是无穷小 ($n \to \infty$ 时). 下定理进一步说明：以概率 1 D_n 是与 $\frac{1}{\sqrt{n}}$ 同级的无穷小.

定理 2 (柯尔莫哥洛夫) 如 $F(x)$ 是 x 的 **连续** 函数，则

$$\lim_{n\to\infty} P\{\sqrt{n} \sup_{-\infty<x<\infty} |F_n(x)-F(x)|<y\}$$
$$= K(y) = \begin{cases} 0, & (y \leqslant 0), \\ \sum\limits_{l=-\infty}^{\infty} (-1)^l e^{-2l^2y^2}, & (y>0). \end{cases} \tag{17}$$

这定理的证明过于复杂，故从略 ([9] 第一版或 [7] 中载有证明).

$K(y)$ 的值有表可查. 采用统计量 $\sqrt{n}\,D_n$ 后，可以像利用皮尔逊定理一样地，利用柯尔莫哥洛夫定理来检验假定 H: "ξ 的分布函数为某 $F(x)$", 注意这里要求 $F(x)$ 已知为 **连续** 函数. 为此，只要对已给信度 $\frac{p}{100}$, 由 $K(y)=1-\frac{p}{100}$ 解出 $y=y_p$; 然后根据 (17), 当 n 充分大时，可认为

$$P\{\sqrt{n}\,D_n<y\}\approx K(y). \tag{18}$$

以子样 $(\xi_1,\cdots,\xi_n)$ 代入 $\sqrt{n}\,D_n$ 后，如 $\sqrt{n}\,D_n<y_p$, 则相对信度 $\frac{p}{100}$ 而言，可接受 H; 如 $\sqrt{n}\,D_n\geqslant y_p$, 则否定 H.

例 6 某工厂生产长度为 2 寸的铁钉，由于生产条件不同，各钉的真实长度 ξ 也不同. 今对已给信度 $\frac{1}{100}$, 欲检验 H : "ξ 有 $N(2,\frac{1}{50})$ 分布". 取出 $n=80$ 钉子作为子样，其长度排成 $\xi_1^*\leqslant\xi_2^*\leqslant\cdots\leqslant\xi_{80}^*$, 从而得经验分布函数 $F_{80}(x)$. 由于正态分布函数 $F(x)$ 连续，故可用定理 2. 设由此子样算得

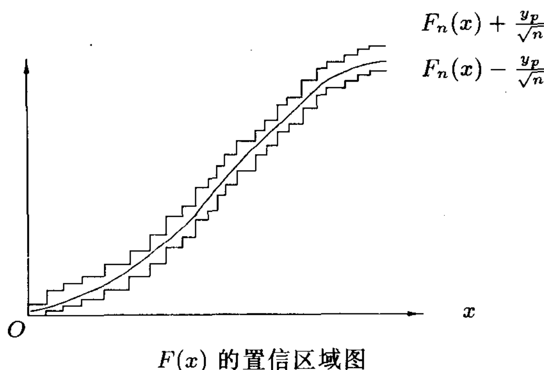
$$\sqrt{80}\cdot D_n=\sqrt{80}\times 0.21=1.88,$$

查分布表得 $y_1=1.63$, 既然 $1.88>1.63$, 故应否定 H.

(17) 式除可用作检验法外，还可用来估计未知分布函数 $F(x)$. 实际上，对已给信度 $\frac{p}{100}$, 仿上取定 y_p. 由 (18) 得知：当 n 充分大时，以近于 $1-\frac{p}{100}$ 的概率，有 $\sqrt{n}\,D_n<y_p$; 亦即对一切 $x\in R_1$, 有

$$F_n(x)-\frac{y_p}{\sqrt{n}}<F(x)<F_n(x)+\frac{y_p}{\sqrt{n}}. \tag{19}$$

这说明，当 n 充分大时，以 $1-\frac{p}{100}$ 左右的概率，$F(x)$ 的图形，**完全** 被包含在 $F_n(x)-\frac{y_p}{\sqrt{n}}$ 与 $F_n(x)+\frac{y_p}{\sqrt{n}}$ 所围成的区域内. 这区域构成 $F(x)$ 的置信区域，置信系数约为 $1-\frac{p}{100}$.(见附图)



$F(x)$ 的置信区域图

*§5.6 最 佳 检 验

(一) 检验法与临界域. 在 §5.5(一) 中，我们提出了判别检验法好坏的标准问题. 在本节中，我们来对它作一个初步的研究.

设 H 是关于随机变数 ξ 的某一假设，根据 ξ 的子样 $\xi_1, \cdots, \xi_n$, 信度 α, 需要判断 H 是否正确. 在 §5.5(一) 中已指出，为了要达到这一目的，构造一个如下形状的、联系于子样的事件 A,

$$A = (g(\xi_1, \cdots, \xi_n) \in B), \tag{1}$$

其中 $g(x_1, \cdots, x_n)$ 是某 n 元波雷耳可测函数，而 B 是某个一维波雷耳集，使在 H 正确的条件下，有

$$P(A) \leqslant \alpha; \tag{2}$$

因而当 H 正确时，A 是一小概率事件，如果 α 充分小的话. 于是当 A 出现时，应该否定 H; 否则就接受 H, 这时犯第一类错误 (弃真) 的概率不大于 α.

由此可见：每一个形如 (1) 而且满足 (2) 的事件 A 决定一种检验法. 一般地，这样的 A 有无穷多个，因为 (1) 中的 g 及 B 可以有无穷多种取法.

例 1 设 ξ 有 $N(a, \sigma)$ 正态分布，a 及 σ 都未知；现在要检验假设 H: $a = 0$. 信度 $\alpha = \frac{p}{100}, (p > 0)$. 在 §5.5 中已知，(1) 中的 A 可取为

$$A = \left(\left|\sqrt{n-1}\ \frac{\overline{\xi}}{s}\right| > t_p\right) \tag{3}$$

(参看 §5.5(3) 式), 因而 (1) 中的

$$B = (-\infty, -t_p) \cup (t_p, \infty). \tag{4}$$

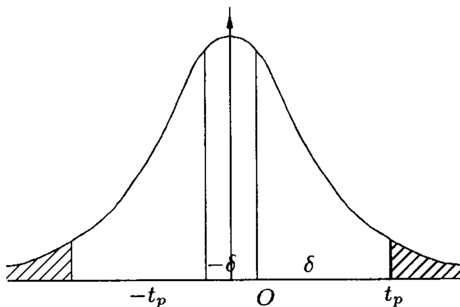


图 1 (其中曲线是 $t(n-1)$ 分布的密度函数的图形)

注意，当 H 正确时，$\sqrt{n-1}\ \frac{\overline{\xi}}{s}$ 有分布 $t(n-1)$; 故 $P(A)$ 等于图 1 中曲线在 $(-\infty, -t_p) \cup (t_p, \infty)$ 上的面积，亦即等于 $\frac{p}{100}$. (3) 中的 A 决定一检验法，今如另选一常数 $\delta > 0$, 使此曲线与 $(-\delta, \delta)$ 围成的面积也等于 $\frac{p}{100}$, 那么

$$A_\delta = \left(\left|\sqrt{n-1}\ \frac{\overline{\xi}}{s}\right| < \delta\right) \tag{5}$$

也满足 (1)(2), 从而也决定一检验法，一般地，在 x 轴上任取波雷耳集 C, 使图 1 中曲线与 C 围成的面积不大于 $\frac{p}{100}$, 则 C 就产生一检验法. 由此可见，我们甚至只用了一个 $g(\xi_1,\cdots,\xi_n)=\sqrt{n-1}\ \frac{\bar{\xi}}{s}$, 就造出了无穷多种检验法.

把 (1) 中的事件 A 改写为

$$A=((\xi_1,\cdots,\xi_n)\in D), \tag{6}$$

其中 $D(\subset R_n)$ 是一 n 维波雷耳集:

$$D=((x_1,\cdots,x_n):g(x_1,\cdots,x_n)\in B). \tag{7}$$

称 D 为临界域，并称由 A 决定的检验法为 D- 检验. D 的作用是：当子样 $(\xi_1,\cdots,\xi_n)\in D$ 时，应否定 H, 否则肯定 H, 信度是 α.

反之，设 D 是任一 n 维波雷耳集，如果在 H 正确的条件下，由此 D 及 (6) 定义的事件 A 的概率不大于 α, 则 A 决定一个检验法，信度是 α, 临界域是 D.

由此可见，为了要给出一检验法，只要给出它的临界域就够了.

(二) **检验法好坏的标准**. 既然检验法一般不唯一，我们自然希望选用一种最好的检验法. 那么所谓“好”的标准是什么？

比较原则 设甲，乙是假设 H 的两种检验法，它们犯第一类错误的概率相等，同为 α, 如果甲犯第二类错误 (取伪) 的概率不大于乙犯第二类错误的概率，我们就说甲比乙好；如果甲比任何一种其它如此的检验法都好，就说甲是具有水平为 α 的最佳检验.

进一步的问题是：最佳检验是否存在？如何找到最佳检验？这依赖于所需要检验的 H 是否复杂. 在一种简单情形，下面的定理 1 完满地解决了这两个问题.

设 ξ 是一连续型随机变数，密度函数 $f(x;a)$ 的形状已知，但含一未知参数 a, 现在有两个假设 H_0,H_1:

$$H_0:\ a=a_0;\quad H_1:\ a=a_1. \tag{8}$$

这里 a_0,a_1 是两个不同的常数. 如果肯定 H_0 就否定 H_1; 如果否定 H_0, 就肯定 H_1. 所以表面上要考虑两个假设，实际上只要检验 H_0 就够了. 称 H_0 为原假设，而 H_1 为备选假设.

取 ξ 的一个子样 $\xi_1,\cdots,\xi_n$, 它们是 n 个相互独立的随机变数，有联合密度为

$$g(x_1,\cdots,x_n;a)=\prod_{i=1}^{n}f(x_i;a). \tag{9}$$

以下用 $P_0(B)$ 表在 H_0 正确的条件下，亦即在 $a=a_0$ 的条件下，事件 B 的概率： $P_0(B)=P(B|H_0)$. 同样 $P_1(B)=P(B|H_1)$, 又为简单起见，记

$$y=(x_1,\cdots,x_n),\quad \eta=(\xi_1,\cdots,\xi_n), \tag{10}$$

y 是一 n 维点， η 是 n 维随机向量. 于是 (9) 中的 $g(x_1,\cdots,x_n;a)$ 可简记为 $g(y;a)$.

对任一非负常数 $c\geqslant 0$, 定义 n 维点集 A_c

$$A_c=(y:\ g(y;a_1)\geqslant cg(y;a_0)); \tag{11}$$

并称商 $\frac{g(y;a_1)}{g(y;a_0)}$ 为似然比. 此外，令

$$\psi(c) \equiv P_0(\eta \in A_c) \equiv P_0\Big(\frac{g(\eta;a_1)}{g(\eta;a_0)} \geqslant c\Big). \tag{12}$$

引理 1 $\psi(c)$ 是 $c \geqslant 0$ 的左连续、单调不增函数，而且

$$0 \leqslant \psi(c) \leqslant \frac{1}{c} \quad (c > 0) \tag{13}$$

$$\psi(0) = 1, \quad \lim_{c\to\infty} \psi(c) = 0. \tag{14}$$

证 由 (12) 知 $\psi(c) \geqslant 0$. 因 $g(y;a_i) \geqslant 0$, 故 $A_0 = R_n$ 而 $\psi(0) = P_0(\eta \in R_n) = 1$. 又对 $c > 0$ 有

$$\begin{aligned}\psi(c) = P_0(\eta \in A_c) &= \int_{A_c} g(y;a_0)\,dy \\ &\leqslant \frac{1}{c}\int_{A_c} g(y;a_1)\,dy \leqslant \frac{1}{c},\end{aligned}$$

(注意这里的积分是 n 重的), 这得证 (13). (14) 中后式自 (13) 推出. 根据 $\psi(c)$ 的定义，并仿照 §2.1 定理 1 的证明，可见 $\psi(c)$ 左连续而且单调不增. □

定理 1 (诺曼 - 皮尔逊 [Neyman-Pearson]) 设随机变数 ξ 有密度函数 $f(x;a)$, 其中 a 为未知参数，又 (8) 式给出原假设 H_0 及备选假设 H_1, 则对任一常数 α, $0 < \alpha < 1$, 必存在水平为 a 的最佳 D- 检验.

证 1° 先设存在常数 c_α, 满足

$$\psi(c_\alpha) = \alpha. \tag{15}$$

在这条件下，我们来证明本定理，然后再解除此条件. 简记

$$P_i(\eta \in A) = F_i(A), \quad (i = 0, 1); \ A_{c_\alpha} = D. \tag{16}$$

下证 (16) 中的 D 决定最佳 D- 检验. 实际上，由 (12)(15)

$$F_0(D) = \psi(c_\alpha) = \alpha. \tag{17}$$

考虑任一 E- 检验，它的临界域是 E, 犯第一类错误的概率等于 α, 因而

$$F_0(E) = \alpha. \tag{18}$$

如果能够证明

$$F_1(D) \geqslant F_1(E), \tag{19}$$

则因犯第二类错误的概率对 D- 检验是 $F_1(\overline{D}) = 1 - F_1(D)$, $\overline{D}$ 表 R_1/D, 对 E- 检验是 $F_1(\overline{E}) = 1 - F_1(E)$, 于是由 (19) 得 $F_1(\overline{D}) \leqslant F_1(\overline{E})$ 而定理结论得以证明.

由 (17)(18) 得

$$F_0(D - ED) = \alpha - F_0(ED) = F_0(E - ED).$$

由 $D=A_{c_\alpha}$ 及 A_{c_α} 的定义，可见如 $y\overline{\in}D$, 则 $c_\alpha g(y;a_0)>g(y;a_1)$. 故

$$\begin{aligned}F_1(D-ED)&\geqslant c_\alpha F_0(D-ED)\\&=c_\alpha F_0(E-ED)\geqslant F_1(E-ED).\end{aligned}$$

在此式两边加上 $F_1(ED)$ 后就得 (19).

2° 任取一满足 $\psi(c_\alpha-0)\geqslant\alpha,\psi(c_\alpha+0)\leqslant\alpha$ 的 c_α, 由引理 1 它必存在. 令 $G=(y:\ g(y;a_1)=c_\alpha g(y;a_0))$, 则

$$\psi(c_\alpha-0)-\psi(c_\alpha+0)=F_0(G)=\int_G g(y;a_0)\,dy.$$

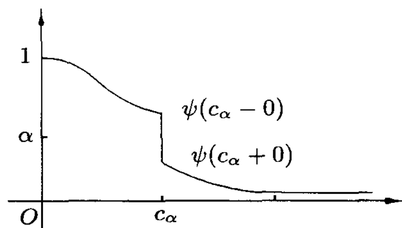


图 2

如果 $\psi(c_\alpha-0)-\psi(c_\alpha+0)=0$, 那么 (15) 成立而定理已证明. 故可设 $\psi(c_\alpha-0)-\psi(c_\alpha+0)>0$, 不妨设 $\psi(c_\alpha-0)-\alpha>0$ [1] 注意 $\psi(c_\alpha-0)-\alpha\leqslant\psi(c_\alpha-0)-\psi(c_\alpha+0)$, 由于分布 F_0 是连续型的，必存在 G 的子集 B, 使 [2]

$$\int_B g(y;a_0)\,dy=\psi(c_\alpha-0)-\alpha. \tag{20}$$

取 $D=(y:g(y;a_1)\geqslant c_\alpha g(y:a_0))-B$, 则由 ψ 的左连续性,

$$\begin{aligned}F_0(D)&=\psi(c_\alpha)-\int_B g(y;a_0)\,dy\\&=\psi(c_\alpha)-\psi(c_\alpha-0)+\alpha=\alpha,\end{aligned}$$

于是重新得到 (17) 式，然后逐字重复 (17) 以下的证明，即知由此 D 决定的 D- 检验法是最佳的. □

由 (11) 可见，当原假设 H_0 中的 a_0 及 c 固定时， A_c 一般依赖于备选假设 H_1 中的 a_1, 因而最佳检验的临界域 D 一般地也依赖于 a_1. 如果当 a_1 在某参数集 Q 中任意变动时， D 与 a_1 无关，我们就说在一切水平为 α 的检验法中，关于 Q 来说， D- 检验是一致的最佳检验. 换句话说，这时不论备选假设 H_1: $a=a_1$ 如何在 Q 中变动，同一临界域都给出最佳检验.

例 2 设 ξ 有 $N(a,\sigma)$ 分布， σ 已知而 a 未知. H_0 为 $a=0, H_1$ 为 $a=a_1$. 对已给水平 $\alpha, 0<\alpha<1$, 我们来求最佳检验的临界域 D. 这时 (11) 右方化为

$$\frac{g(y;a_1)}{g(y;0)}=e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}[(x_i-a_1)^2-x_i^2]}=e^{Hz-\frac{1}{2}H^2}\geqslant c, \tag{21}$$

[1] 如果 $\psi(c_\alpha-0)-\alpha=0$, 则由引理 1 中的左连续性， $\psi(c_\alpha)=\alpha$ 而化为 1° 中情形.

[2] 实际上，令 $G_r=G\cap(-\infty,r]$, $(-\infty,r]$ 表 n 维区间， $r=(r_1,\cdots,r_n)$, r_i 为有理数. 再令 $s=\sup(r:F_0(G_r)\leqslant\psi(c_\alpha-0)-\alpha)$, 则可取 $B=G\cap[-\infty,s]$. 因此

$$\int_B g(y;a_0)\,dy=\lim_{r\uparrow s}\int_{G_r} g(y;a_0)\,dy=\lim_{r\uparrow s}F_0(G_r)=\psi(c_\alpha-0)-\alpha.$$

后一等号是由于连续函数在闭区间内必取上确界

其中 $H=\sqrt{n}\ \frac{a_1}{\sigma}$ 是常数， $z=\sqrt{n}\ \frac{\overline{x}}{\sigma}$, $\left(\overline{x}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i\right)$. 因此,

$$\psi(c)=P_0\left(e^{H\cdot\sqrt{n}\ \frac{\overline{\xi}}{\sigma}-\frac{1}{2}H^2}\geqslant c\right). \tag{22}$$

先设 $a_1>0$, 从而 $H>0$. 取 b 使 $\frac{1}{\sqrt{2\pi}}\int_b^{\infty}e^{-\frac{x^2}{2}}dx=\alpha$, 我们证明,

$$c_{\alpha}=e^{Hb-\frac{1}{2}H^2} \tag{23}$$

满足 (15), 实际上，因 $\overline{\xi}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 的分布在 H_0 正确的条件下为 $N\left(0,\frac{\sigma}{\sqrt{n}}\right)$, 故由 (22) 得

$$\psi\left(e^{Hb-\frac{1}{2}H^2}\right)=P_0\left(\sqrt{n}\ \frac{\overline{\xi}}{\sigma}\geqslant b\right)=\frac{1}{\sqrt{2\pi}}\int_b^{\infty}e^{-\frac{x^2}{2}}dx=\alpha.$$

现在以 (23) 中的 c_{α} 代入 (21) 中的 c, 可见所求的临界域为

$$D=\left(y:\sqrt{n}\ \frac{\overline{x}}{\sigma}\geqslant b\right). \tag{24}$$

由于 D 与 a_1 无关，故关于 $Q=(0,\infty)$ 而言，它还决定水平为 α 的一致最佳检验.

次设 $a_1<0$. 同样求得关于 $Q'=(-\infty,0)$ 的水平为 α 的一致最佳检验的临界域为 D':

$$D'=\left(y:\sqrt{n}\ \frac{\overline{x}}{\sigma}\leqslant -b\right). \tag{25}$$

*§5.7 若 干 应 用

数理统计在工业、农业、医药卫生等许多实际部门有着广泛的应用，在本书中，我们不可能作详细的叙述，只好作一简单的介绍，目的是想说明本章的理论，可以用来解决那些实际问题以及解题的基本思想. 至于详细的情形，请看书末所引的参考书 [3], [31],[32],[33],[36].

(一) 方差分析. 方差分析是一种统计分析的方法，首先由 R. A. Fisher 所引进，它有许多应用，例如，可用以比较各种植物品种，动物品种，各种原料或各种实验方法的优缺点等等.

考虑下列问题：设有 r 个独立，正态分布的随机变数 $\xi_1,\xi_2,\cdots,\xi_r$, 它们有相同的未知方差 σ^2, 但未知的数学期望 $a_1,a_2,\cdots,a_r$ 可能不同. 现在对每个随机变数 ξ_i 作一组独立的观察而得子样

$$\xi_{i1},\xi_{i2},\cdots,\xi_{in_i},\quad (i=1,\cdots,r),$$

试根据这 r 组观察来检查假设 H

$$H:\ a_1=a_2=\cdots=a_r.$$

例如， $\xi_1,\xi_2,\cdots,\xi_r$ 可以是 r 种不同品种的小麦的亩产量，或者是用 r 种不同方法所生产的电灯泡的使用时数等等. 如果 H 不正确，这说明 r 种小麦的平均亩产量不相同，也就是说这些品种是有差异的，于是进一步我们可挑选优良的品种来提高产量.

当 $r=2$ 时，这问题已在 §5.5(二)3° 中研究过，一般情况的解法如下. 令

$$\bar{\xi}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\xi_{ij}, \quad n = \sum_{i=1}^{r} n_i, \tag{1}$$

$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\xi_{ij} = \frac{1}{n}\sum_{i=1}^{r} n_i\bar{\xi}_i \tag{2}$$

因而 $\bar{\xi}_i$ 及 $\bar{\xi}$ 分别是第 i 组子样及全体子样的平均值. 容易证明恒等式

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi})^2 = \sum_{i=1}^{r} n_i(\bar{\xi}_i-\bar{\xi})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi}_i)^2; \tag{3}$$

简记 (3) 为

$$Q = Q_1 + Q_2 \tag{4}$$

Q 是 ξ_{ij} 对全体子样总平均值 $\bar{\xi}$ 的离差平方和， Q_1 是第 i 组子样的平均值对总平均值的离差平方和，而 Q_2 是每组子样 ξ_{ij} 对本组平均值的离差平方和的和. (4) 式把 Q 拆成两部分. 注意 Q/n 是全体子样的方差.

设 H 正确，由 §5.2 定理 1,Q/σ^2 有 $\chi^2(n-1)$ 分布，与 §5.2 定理 1 的证明相似，可证明： Q_1/σ^2 及 Q_2/σ^2 分别有 $\chi^2(r-1)$ 及 $\chi^2(n-r)$ 分布，而且相互独立. 根据 §2.8 定理 4, 可见随机变数

$$\zeta = \frac{Q_1}{r-1}\Big/\frac{Q_2}{n-r} \tag{5}$$

具有自由度为 $r-1$ 及 $n-r$ 的 F 分布. 然后利用以前多次用过的方法，对已给信度 $p\%$, $0<p<100$, 可找到常数 f_p, 使

$$P(\zeta > f_p) = p\%.$$

于是判别 H 的准则是：对信度 $p\%$ 而言，如果 $\zeta \leqslant f_p$, 则接受 H, 否则拒绝 H.

如果检验结果发现 H 不正确，我们进一步还可利用 §5.4(二)3° 中的方法来估计差 $a_i - a_j$, $(1\leqslant i,j\leqslant r)$.

以上所考虑的差异是由于一种原因 (例如在影响小麦亩产量的问题中只考虑品种一个因素) 引起的，所以叫做一元方差分析. 自然还可考虑多种因素 (例如，除品种外，还有肥料，土质，水分，阳光等等因素), 这便产生多元方差分析. 解题的基本思想与一元时类似，关键在于把 (4) 式扩充，例如考虑两个因素时，应扩充为

$$Q = Q_1 + Q_2 + Q_3.$$

(二) 回归分析. 我们仍旧考虑上面的实例. 小麦的亩产量 ξ 与所施肥料量 z 有关. 在一定范围内，如果施肥较多，即 z 较大时，亩产量 ξ 也较大，因而 ξ 依赖于 z, 我们希望知道 ξ 是怎样依赖于 z 的. 首先注意，当 z 固定时，通常可假定 ξ 有正态分布 $N(a,\sigma)$ (根据中心极限定理及长期的实践经验), 方差 σ^2 与 z 无关，但平均值

$$a = a(z) \tag{6}$$

是 z 的函数，于是问题化为要研究 $a(z)$: 或者根据历年的经验对 $a(z)$ 作某假设 H(譬如，H: $a(z)$ 是 z 的线性函数，即 $a=u+vz$, u,v 是二常数), 并检验 H 是否正确，或者估计 $a(z)$.

于是我们来作一些实验，分别给 z 以 n 个不同的值 z_i, 相应地得到 n 个观察值 ξ_i; 换句话说，施肥量为 z_i 时，得亩产量为 ξ_i, 从而得一组观察值 (子样) 为

$$(z_1,\xi_1),(z_2,\xi_2),\cdots,(z_n,\xi_n),$$

这就是我们进行回归分析的数据.

有了这个实例，容易理解下列回归问题的一般提法：设 ξ 是有正态分布 $N(a,\sigma)$ 的随机变数，其中

$$a=a(z^{(1)},z^{(2)},\cdots,z^{(k)}) \tag{7}$$

是非随机的变量 $z^{(1)},z^{(2)},\cdots,z^{(k)}$ 的函数， σ 虽未知，但与这些变量无关；试根据一系列的观察

$$(z_1^{(1)},z_1^{(2)},\cdots,z_1^{(k)},\xi_1),\cdots,(z_n^{(1)},z_n^{(2)},\cdots,z_n^{(k)},\xi_n),$$

其中 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立的随机变数，来对 a 作统计检验或估计.

为了说明解法的基本思想，我们只考虑一个变量 z 而且 H 为

$$a=\alpha'+\beta z$$

的简单情形，这里 α',β 是未知常数，引进 $\alpha=\alpha'+\beta\overline{z}$, $\overline{z}=\frac{1}{n}\sum\limits_{i=1}^{n}z_i$, 为了研究 a',β, 只要研究 α,β 就够了. 改写上式为

$$a=\alpha+\beta(z-\overline{z}). \tag{8}$$

先求 α,β 及 σ 的估值，用最大似然法 (比较 §5.3 例 2). 注意 $\xi_1,\xi_2,\cdots,\xi_n$ 的联合密度是

$$f(x_1,x_2,\cdots,x_n)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[x_i-\alpha-\beta(z_i-\overline{z})]^2\right\}.$$

用最大似然法求得 α,β,σ 的估值分别为

$$\alpha^*=\overline{\xi},\quad \beta^*=\frac{\sum\limits_{i=1}^{n}(z_i-\overline{z})(\xi_i-\overline{\xi})}{\sum\limits_{i=1}^{n}(z_i-\overline{z})^2}=\frac{m_{11}}{s_1^2}, \tag{9}$$

$$\sigma^{*2}=\frac{1}{n}\sum_{i=1}^{n}[\xi_i-\alpha^*-\beta^*(z_i-\overline{z})]^2=s_2^2(1-r^2); \tag{10}$$

其中 $s_1^2=\frac{1}{n}\sum\limits_{i=1}^{n}(z_i-\overline{z})^2$, $s_2^2=\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\overline{\xi})^2$, 其余的 m_{11} 及 r^2 容易算出.

其次求 α^*,β^* 及 σ^{*2} 的分布. 由 (9) 知 α^*,β^* 都是 ξ_i 的线性函数，故它们都有正态分布，于是只要求出它们平均值与方差就够了：

$$E\alpha^*=\alpha,\quad D\alpha^*=\frac{\sigma^2}{n}, \tag{11}$$

$$E\beta^*=\beta,\quad D\beta^*=\frac{\sigma^2}{ns_1^2}. \tag{12}$$

我们有恒等式

$$\sum_{i=1}^{n}[\xi_i-\alpha^*-\beta^*(z_i-\overline{z})]^2$$
$$=\sum_{i=1}^{n}[\xi_i-\alpha-\beta(z_i-\overline{z})]^2-n(\alpha^*-\alpha)^2-ns_1^2(\beta^*-\beta)^2. \tag{13}$$

引进随机变数 $\eta_1,\cdots,\eta_n$:

$$\eta_i=\xi_i-\alpha-\beta(z_i-\overline{z}),$$

显然它们相互独立，各有 $N(0,\sigma)$ 分布. 下面二线性型

$$\left.\begin{aligned}\zeta_1&=\sqrt{n}\,(\alpha^*-\alpha)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i,\\ \zeta_2&=\sqrt{n}\,s_1(\beta^*-\beta)=\frac{1}{\sqrt{n}\,s_1}\sum_{i=1}^{n}(z_i-\overline{z})\eta_i\end{aligned}\right\} \tag{14}$$

满足正交条件:

$$\left.\begin{aligned}&\left(\frac{1}{\sqrt{n}}\right)^2+\cdots+\left(\frac{1}{\sqrt{n}}\right)^2=1,\\ &\frac{1}{ns_1^2}\sum_{i=1}^{n}(z_i-\overline{z})^2=1,\\ &\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\ \frac{1}{\sqrt{n}\,s_1}(z_i-\overline{z})=0.\end{aligned}\right\} \tag{15}$$

由于 (15), 可以找到正交矩阵 (c_{ij}), $i,j=1,2,\cdots,n$, 使

$$\zeta_i=\sum_{j=1}^{n}c_{ij}\eta_j,\quad (i=1,2,\cdots,n), \tag{16}$$

而且 (16) 中当 $i=1,2$ 时的两个式子与 (14) 一致;因而经过正交变换 (c_{ij}), 我们把 $(\eta_1,\cdots,\eta_n)$ 变到 $(\zeta_1,\cdots,\zeta_n)$. 由于

$$E\zeta_i\zeta_j=\sigma^2\sum_{k=1}^{n}c_{ik}c_{kj}=\begin{cases}\sigma^2, & i=j,\\ 0, & i\neq j,\end{cases}$$

可见 $\zeta_1,\cdots,\zeta_n$ 是相互独立的随机变数，各有正态分布 $N(0,\sigma)$. 由 (10) (13) 及 η_i 的定义得

$$\begin{aligned}n\sigma^{*2}&=\sum_{i=1}^{n}\eta_i^2-\zeta_1^2-\zeta_2^2\\ &=\sum_{i=1}^{n}\zeta_i^2-\zeta_1^2-\zeta_2^2=\sum_{i=3}^{n}\zeta_i^2.\end{aligned}$$

故由 §2.8 定理 2 知 $\frac{n\sigma^{*2}}{\sigma^2}$ 有 $\chi^2(n-2)$ 分布.

第三，既然 $\alpha^*,\beta^*,\sigma^{*2}$ 的分布已知，关于 α,β,σ^2 的区间估值和假设检验的问题就迎刃而解. 关于 σ^2 只要仿照 §5.4(二)2° 及 §5.5(二)2°, 统计量改用 $\frac{n\sigma^{*2}}{\sigma^2}$. 由 §2.8 定理 3 及上面的结果，可见二随机变数

$$\sqrt{n-2}\ \frac{\alpha^*-\alpha}{\sigma^*}\ \text{及}\, s_1\sqrt{n-2}\ \frac{\beta^*-\beta}{\sigma^*}$$

各有 $t(n-2)$ 分布. 注意 s_1 已知，故关于 α 及 β 的区间估值和假设检验问题完全可仿照 §5.4(二)1° 及 §5.5(二)1° 来解决.

最后来叙述预测问题，它在实际中很有用. 设已给定 z 的值，试求 a 的置信区间，置信系数为 $1-\frac{p}{100},(0<p<100)$. 用上面的实例来解释一下：我们打算施肥 z (公斤), 试预测今年的平均产量 a; 或者，更精确些，试找到一个 (随机) 区间 $(\widehat{a}_1,\widehat{a}_2)$, 使

$$P(\widehat{a}_1<a<\widehat{a}_2)=1-\frac{p}{100}.$$

为此，由 (11)(12) 并注意 ζ_1,ζ_2(从而 α^*,β^*) 的独立性，知

$$X=\frac{\alpha^*+\beta^*(z-\overline{z})-a}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{n^2s_1^2}(z-\overline{z})^2}}$$

有 $N(0,1)$ 分布，既然 $\frac{n\sigma^{*2}}{\sigma^2}$ 有 $\chi^2(n-2)$ 分布，故由 §2.8 定理 3, 可见

$$t=X\Bigg/\sqrt{\frac{n\sigma^{*2}}{(n-2)\sigma^2}}=\frac{\sqrt{n-2}}{\sqrt{1+\left(\frac{z-\overline{z}}{s_1}\right)^2}}\cdot\frac{\alpha^*+\beta^*(z-\overline{z})-a}{\sigma^*}\tag{17}$$

有 $t(n-2)$ 分布，故置信区间的两个端点 $\widehat{a}_1,\widehat{a}_2$ 为

$$\alpha^*+\beta^*(z-\overline{z})\pm t_p\frac{\sigma^*}{\sqrt{n-2}}\sqrt{1+\left(\frac{z-\overline{z}}{s_1}\right)^2},\tag{18}$$

其中常数 t_p 可由 $t(n-2)$ 分布表找到.

(三) **质量控制**. 在生产过程中常常采用统计方法，以检查和控制产品的质量. 这种方法既可用于生产过程中的产品，也可用于已制成的产品 (成品). 我们先从前一情形讲起.

(A) 生产过程一般是稳定的，就是说，如果影响质量的主要因素如原料、工序、机器、人力等无显著变化时，那么产品质量也不会发生重大差异. 譬如说，生产出的钢条的直径，不会有重大改变，虽然如此，即使生产过程是稳定的，影响质量的主要因素还是会有随机的波动，不过波动较小罢了. 现在假定希望制造直径是 a 厘米的钢条，由于这些波动，制成的钢条的直径 ξ 是一随机变数. 在假设 H: "生产过程稳定" 的条件下，可以近似地假定 (根据中心极限定理): ξ 有 $N(a,\sigma)$ 正态分布.

运用统计的方法，可以在废品尚未出现以前，就能预告生产稳定性即将破坏，事故即将出现，并且能在一定程度上指出毛病所在，以便生产者及时采取行动，防止废品的出现. 这种质量控制是积极性的，我们称它为预告检查.

预告检查的基本思想如下：在假设 H: "生产过程稳定" 的条件下，某产品的一个主要数字特征 ξ(例如钢条的直径) 可以认为有 $N(a,\sigma)$ 分布 (或近似 $N(a,\sigma)$ 分布), 对已给信度 $\alpha, 0<\alpha<1$, 由正态分布表，存在常数 n_α, 使

$$P(-n_\alpha\sigma<\xi-a<n_\alpha\sigma)=1-\alpha;\tag{19}$$

因而

$$|\xi-a|\geqslant n_\alpha\sigma \tag{20}$$

是一小概率 (概率为 α) 事件. 现在对一列产品进行观察，得子样

$$\xi_1,\xi_2,\cdots,\xi_n, \tag{21}$$

以 ξ_1 代替 (20) 中的 ξ 后，如果 (20) 不成立，这说明在制成第一个抽样产品时 H 正确，即生产稳定 (信度为 α), 于是同样考虑 $\xi_2,\cdots$, 如果对某 ξ_k, (20) 成立，则在制成第 k 个抽样产品时应否定 H, 就是说，这时生产的稳定性已被破坏 (信度 α). 于是应采取修理行动.

一般地，常数 a 是预定的， σ^2 可根据工厂以往的记录找到，或以 $s^2=\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\bar{\xi})^2$ 作为它的估值.

通常取 $1-\alpha=95.5\%$ 或 99.7%, 这时 n_α 等于 2 或 3.

更直观地，可以制一控制图 (图 1): 设 $1-\alpha=99.7\%$, 以 $y=a$ 为中心线，再作四直线

$$L_1:\ y=a+3\sigma,$$
$$L_2:\ y=a-3\sigma,$$
$$M_1:\ y=a+2\sigma,$$
$$M_2:\ y=a-2\sigma.$$

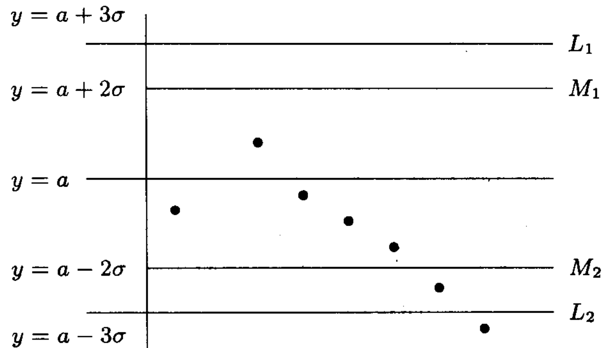


图 1

称 L_1(或 L_2) 为上 (或下) 废品线，称 M_1(或 M_2) 为上 (或下) 控制线，当 ξ_k 跑出 L_1 与 L_2 所围成的地带时，生产已不稳定；当 ξ_k 跑到 M_1 之上或 M_2 之下时，警告着生产稳定有破坏的危险，因而 M_1,M_2 起着警告作用. 图 1 在 $k=6$ 时提出警告，在 $k=7$ 时出现废品.

利用控制图，不必等到点落在控制区外，往往就能看出问题. 例如，第一，如点总是落在中心线以上 (或以下), 可能是机器安装得不适当，或机件已受损耗，或零件安装太紧 (太松) 等；第二，点在控制区内波动很大，一下靠近上控制线，一下又靠近下控制线，这可能是某一因素 (如原料) 有较大波动，或发动机有毛病；第三，如控制图中的点的联线有某规律 (如周期现象), 那么可能是机器有系统的故障 (如周期性的故障); 最后，如点有单调下降 (见图 1) 或上升，可能是工具磨损或触媒劣化等原因.

有时候，如 ξ 的分布不很接近正态分布，则上法的效果要降低. 这时可改用

$$g(\xi_1,\cdots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i \tag{22}$$

来代替 ξ, 我们知道，由 §3.4 定理 1,$\bar{\xi}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 的分布近似于 $N\big(a,\frac{\sigma}{\sqrt{n}}\big)$. 在采用 $\bar{\xi}$ 进行控制时，应观察一列子样

$$\xi_{i1},\xi_{i2},\cdots,\xi_{in}\quad(i=1,2,\cdots).$$

除了 $\bar{\xi}$ 外，有时还采用子样的散度

$$R = \max(\xi_1, \cdots, \xi_n) - \min(\xi_1, \cdots, \xi_n)$$

来控制，类似地可以作控制图. 如果要求严格，可以用 $\bar{\xi}$ 与 R 作两张控制图，联合控制.

控制图是生产过程的简单摄影，对生产是很有用的.

(B) 现在来叙述成品的质量控制. 设有 N 个产品 (如手表), 验收人员需要检查每个产品是合格 (成品) 还是不合格 (废品). 如果 N 很大，要逐一检查是很难做到的，于是只好从中随机地抽取含 $n(< N)$ 个个体的子样，希望以这 n 个的逐一检查，来代替含 N 个个体的母体的全面检查. 这种检查方法叫抽样检查. 它的优点是，需要检查的个体少了；然而，它是否真能反映母体的质量呢？容易想象，如果 n 越大，就越能保证质量；但另方面，检查的劳动量及费用也就越增加，矛盾的焦点就在于此. 从而可见，如何合理地选取 n, 是一个重要的问题.

母体质理好坏的标志是

$$\text{废品率 } p = \frac{1}{N} \times \text{ 母体中废品个数}.$$

根据国家对质量的要求，常常可以定出两个数 $p_1, p_2, 0 < p_1 < p_2 < 1$, 当 $p \leqslant p_1$ 时，我们说母体合格，当 $p \geqslant p_2$ 时，便认为母体不合格.

验收人员采用下列简单的验收方案：从母体中抽取容量为 n 的子样，子样中的废品数 d 如不大于 r, 他接收母体；如 $d > r$, 便拒收母体. 验收人员自然希望，接收合格母体的概率不小于 $1 - \varepsilon_1$(充分大) 接收不合格母体的概率不大于 ε_2(充分小), $0 < \varepsilon_1, \varepsilon_2 < 1$. 本着这种要求，试问如何定出 n 及 r? 一旦决定了 n 与 r, 验收方案便完全订妥，因此可称它为 (n, r) 方案.

以 $g(p)$ 表废品率为 p 时接收母体的概率，令 $q = 1 - p$. 则

$$g(p) = \sum_{i=0}^{r} \frac{\binom{Np}{i}\binom{Nq}{n-i}}{\binom{N}{n}}, \tag{23}$$

如果相对于 n 来说 N 充分大，根据 §2.3(11),$g(p)$ 近似于 $f(p)$:

$$f(p) = \sum_{i=0}^{r} \binom{n}{i} p^i q^{n-i}.$$

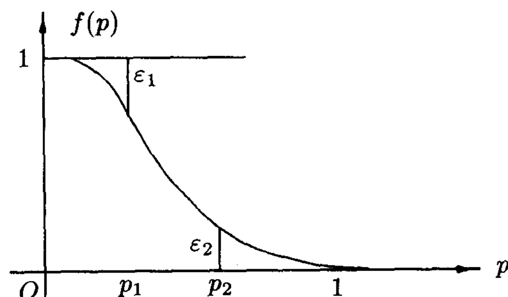


图 2

当 p 自 0 上升到 1 时， $f(p)$ 不增，连续.

$$f(0) = 1, \quad f(1) = 0.$$

(图 2).

根据上述验收人员的要求，得到二方程

$$f(p_1) = 1 - \varepsilon_1, \tag{24}$$

$$f(p_2) = \varepsilon_2. \tag{25}$$

由此二方程可解出 n 及 r，于是方案订妥．执行这一方案，可以保证接受合格母体的概率不小于 $1-\varepsilon_1$，接受不合格母体的概率不大于 ε_2．

我们看到，从验收人员的要求 (他自然应该反映国家的要求) 出发可以确定 n,r; 如果换成另外的要求，自然也可以另订方案，不过我们不去深入叙述了．

一种更完善的检验方法是 A. Wald 所建立的序贯分析，关于这方面的详细叙述可参看他的著作 [36]．

习　题

1. 设 X 在 $[\theta-a,\theta+a]$ 上均匀分布，$(x_1,\cdots,x_n)$ 是一组样本．试证：$\overline{x}$ 与 $\frac{1}{2}\left(\max\limits_{1\leqslant i\leqslant n} x_i+\min\limits_{1\leqslant i\leqslant n} x_i\right)$ 都是 θ 的无偏估值，两者谁较有效？

2. 设 x_1,x_2,x_3 是 X 的三个样本，X 的分布密度是

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta}, & 0<x<\theta,\\ 0, & \text{其他},\end{cases}\qquad 0<\theta<\infty.$$

试证 $\frac{4}{3}\max\limits_{1\leqslant i\leqslant 3} x_i$, $4\min\limits_{1\leqslant i\leqslant 3} x_i$ 都是 θ 的无偏估值，并比较它们谁有效？

3. 设 $\widehat{\theta}_1(x_1,\cdots,x_n)$ 和 $\widehat{\theta}_2(x_1,\cdots,x_n)$ 是参数 θ 的两个独立的无偏估计，并且 θ_1 的方差是 θ_2 方差的两倍，试找出常数 K_1,K_2 使得 $K_1\widehat{\theta}+K_2\widehat{\theta}_2$ 是 θ 的无偏估值，并且在所有这样的线性估值中方差最小．

4. 设 $x_1,\cdots,x_n$ 是一带有参数 θ 的总体的 n 个样本，$t_n(x_1,\cdots,x_n)$ 是 θ 的一个估值，若 $Et_n=\theta+k_n$, $Dt_n=\sigma_n^2$．且

$$\lim_{n\to\infty} k_n=\lim_{n\to\infty}\sigma_n^2=0$$

则 t_n 是 θ 的一致性估值．

5. 设 X 具有分布密度

$$f(x,\theta)=\begin{cases}1, & \theta-\dfrac{1}{2}\leqslant x\leqslant\theta+\dfrac{1}{2}, -\infty<\theta<\infty,\\ 0, & \text{其他}.\end{cases}$$

$x_1,\cdots,x_n$ 是 X 的 n 个样本，试求 θ 的最大似然估值．

6. 设 X 的分布密度是

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta}, & 0<x\leqslant\theta,\ 0<\theta<\infty,\\ 0, & \text{其他}.\end{cases}$$

$x_1,\cdots,x_n$ 是 X 的 n 个样本，试求 θ 的最大似然估值．

7. 设 $x_1,\cdots,x_n$ 是下列分布密度的 n 个样本，试求 θ 的最大似然估值．

(i) $f(x,\theta)=\begin{cases}\dfrac{\theta^x e^{-\theta}}{x!}, & x=0,1,2,\cdots,\ 0\leqslant\theta<\infty,\\ 0, & \text{其他};\end{cases}$

(ii) $f(x,\theta)=\begin{cases}\theta x^{\theta-1}, & 0<x<1,\ 0<\theta<\infty,\\ 0, & \text{其他};\end{cases}$

(iii) $f(x,\theta)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & 0<x<\infty,\ 0<\theta<\infty,\\ 0, & \text{其他};\end{cases}$

(iv) $f(x,\theta)=\frac{1}{2}e^{-|x-\theta|}$, $-\infty<x<\infty$, $-\infty<\theta<\infty$;

(v) $f(x,\theta)=\begin{cases}e^{-(x-\theta)}, & \theta\leqslant x<\infty,\ -\infty<\theta<\infty,\\ 0, & \text{其他};\end{cases}$

8. 设 $x_1,\cdots,x_n$ 是下列分布密度的 n 个样本，试求对参数 θ_1 和 θ_2 的最大似然估值

$$f(x,\theta_1,\theta_2)=\frac{1}{\theta_2}e^{-\frac{x-\theta_1}{\theta_2}},\ \theta_1<x<\infty,\ -\infty<\theta_1<\infty,\ 0<\theta_2<\infty.$$

9. 从甲，乙两店各买同样重量的豆，在甲店买了 13 次，在乙店买了 10 次，每次买得的豆的颗数分别为 $(x_1,\cdots,x_{13})$, $(y_1,\cdots,y_{10})$. 设豆的颗数是服从正态分布，若 $\overline{x}=118$, $\overline{y}=116.1$, $\Sigma(x_i-\overline{x})^2=2825$, $\Sigma(y_i-\overline{y})^2=1442$. 取信度为 1%, 问是否可以认为甲，乙两店的豆是同一种类的 (即同样重量的豆颗数应该一样).

10. 从某种试验物中取来 24 个样品 $(x_1,\cdots,x_{24})$ 测量其发热量，发热量看作是正态分布，若量得 $\overline{x}=11958$, $s_n=323$, 问以 5% 的信度是否可认为发热量的期望值是 12100.

11. 由十块地试种甲，乙两品种作物，所得产量为 $(x_1,\cdots,x_{10})$, $(y_1,\cdots,y_{10})$, 假设作物产量服从正态分布，并计算得 $\overline{x}=30.97$, $\overline{y}=21.79$, $s_x^2=26.7$, $s_y^2=12.1$. 若取信度为 1%, 问是否可以认为这二种品种的产量没有显著性差别.

12. 两台机床加工同一零件，分别取 6 个和 9 个零件，量得其口径为 $(x_1,\cdots,x_6)$, $(y_1,\cdots,y_9)$, 又 $\overline{x}=34.1$, $\overline{y}=41.15$, $s_x^2=0.345$, $s_y^2=0.357$. 若取信度为 5%, 假定零件口径服从正态分布，问是否可以认为两台机床所加工的零件口径的方差无差异.

13. 在某公路上， 50 分钟之间，观测每 15 秒过路的汽车的辆数，得到的次数分布如下

过路的辆数 x	0	1	2	3	4	5
次　数　n_x	92	68	28	11	1	0

问这个分布能否看成是普阿松分布？ (信度取 10%).

*第六章　随机过程的模拟

§6.1　在电子计算机上模拟均匀分布随机变数的方法

(一) 问题的提出．在 §2.8 中，我们讨论了随机向量的变换问题，那里是先给定了一个变换，要求的是经过变换后的随机向量的分布．本章中所要讨论的问题可以在一定意义上看成那里的逆，即我们希望变换后的随机向量具有事先指定的分布，试问如何找到所需的变换？此外，还要求变换比较简单，以便能在电子计算机上较快地实现．

这个问题具有重要的实际意义．第一，在许多工程、通讯等技术问题中，所研究的控制过程往往不可避免地伴有随机因素，要使理论符合实际情况，必须把这些因素考虑在内，这样就自然地会产生随机过程的模拟问题．其次，不久前发展起来了一种有效的计算方法－统计试验计算法 (或名蒙特卡罗 [Monte-Carlo] 法)，这种方法的基本思想是人为地造出一种概率模型，使它的某些参数恰好就是所考虑的问题的解．例如，在 §1.1 中的蒲丰问题里，所要计算的 π 恰好出现在投针试验的概率中．因此，要利用这种计算方法，先决条件是要把所需的概率模型模拟出来．在下章里，我们要比较详细地讨论蒙特卡罗方法．

实际中的问题多种多样，所要模拟的过程是非常广泛的，我们在这里只能介绍一些最基本的结果．读者有进一步的需要时可以参阅文献 [37].

(二) 均匀分布随机变数的产生．先从最简单的问题开始．设已给一单调上升的连续分布函数 $F(x)$, 试作一随机变数，使它的分布函数重合于 $F(x)$. 这个问题有多种解法 (参看 §2.2 定理 2), 为了便于在计算机上实现，这里采用下面的方法．任取一在 (0,1) 中均匀分布的随机变数 ξ, 显然它的分布函数 $G(x)$ 为

$$G(x)=\begin{cases}0, & \text{如} x\leqslant 0,\\ x, & \text{如} x\in(0,1),\\ 1, & \text{如} x\geqslant 1.\end{cases}$$

容易证明， $F^{-1}(\xi)$ 即所求的随机变数：这里 $F^{-1}(x)$ 表 $F(x)$ 的反函数，即 $F(F^{-1}(x))=x$ 事实上

$$P(F^{-1}(\xi)\leqslant x)=P(\xi\leqslant F(x))=G(F(x))=F(x),$$

这证明 $F^{-1}(\xi)$ 的分布函数是 $F(x)$(参看 §2.11 例 4).

注意 $F^{-1}(\xi)$ 是含未知数 y 的方程

$$F(y)=\xi \tag{1}$$

的解；如果 $F(x)$ 有密度为 $f(x)$, 那么 (1) 化为

$$\int_{-\infty}^{y} f(x)dx=\xi, \tag{2}$$

因而 $F^{-1}(\xi)$ 满足 $\int_{-\infty}^{F^{-1}(\xi)} f(x)dx=\xi$. 于是得

定理 1 设已给单调上升的连续分布函数 $F(x)$(或已给分布密度 $f(x)$) 及在 (0,1) 中均匀分布的随机变数 ξ, 则方程 (1)(或 (2)) 的解 $y=F^{-1}(\xi)$ 是以 $F(x)$ 为分布函数的随机变数 (见图 1).

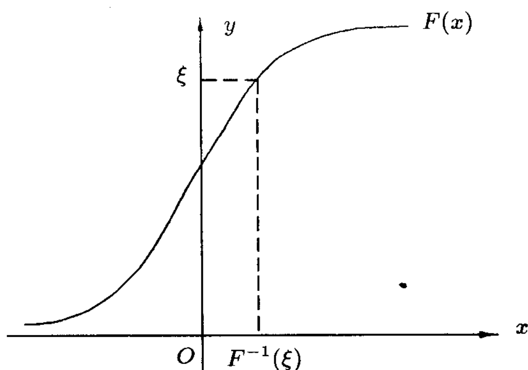


图 1

例 1 设已给指数分布函数为

$F(x)=1-e^{-\lambda x},(x>0),\lambda$ 为正常数.

(1) 化为 $1-e^{-\lambda y}=\xi$,

解得 [1]

$y=-\frac{1}{\lambda}\ln(1-\xi)$.

由定理 1 可见，要造出以上述 $F(x)$ 为分布函数的随机变数，只要会造出 (0,1) 上均匀分布的随机变数就够了. 为方便计，以下简称后者为均匀变数.

为了造出均匀变数，采用二进位记数法. (0,1) 中的任一数可表为

$$\xi=0.\eta_1\eta_2\eta_3\cdots\equiv\frac{\eta_1}{2}+\frac{\eta_2}{2^2}+\frac{\eta_3}{2^3}+\cdots, \tag{3}$$

其中 η_i 或者等于 0, 或者等于 1.

反之，任取一列相互独立的随机变数 $\eta_1,\eta_2,\cdots$, 它们有相同的分布密度矩阵 $\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$, 即 η_i 等可能地取及 0 及 1 为值，则由 (3) 造出的 ξ 在 (0,1) 中均匀分布. 为了证明这个结论，任取一区间 $(a,b),0\leqslant a<b\leqslant 1$, 要证的是 $P(a<\xi<b)=b-a$, 由于 a(或 b) 可由有限小数 (即 (3) 中除有限多个 η_i 外，其余 η_j 全为 0 的数) 所任意逼近，因此只需考虑 a,b 为有限小数的情形，为了突出问题的本质并避免符号上的烦琐，设 $(a,b)=(\frac{1}{4},\frac{3}{4})$. 注意

$$\left(\frac{1}{4},\frac{3}{4}\right)=\left(\frac{1}{4},\frac{1}{2}\right)\bigcup\left[\frac{1}{2},\frac{3}{4}\right),$$

采用二进制，后面二区间可写成

$$(0.01,0.01111\cdots)\cup[0.10,0.1011\cdots),$$

由此可见：要使 ξ 取值于前 (或后) 一区间，充要条件是 $\eta_1=0,\eta_2=1$(或$\eta_1=1,\eta_2=0$). 因此

$$P\left(\frac{1}{4}<\xi<\frac{3}{4}\right)=P(\eta_1=0,\eta_2=1)+P(\eta_1=1,\eta_2=0)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}=\frac{3}{4}-\frac{1}{4}.$$

对一般的 (a,b) 可类似证明.

这样一来，造均匀变数的问题化为造一列独立的 $\{\eta_i\}$ 的问题，而 $\{\eta_i\}$ 的造法多种多样. 譬如说，连续不断地独立地掷硬币而记录其正面为 1, 反面为 0. 或者在一段确定的时

[1] 由于 $1-\xi$ 也在 (0,1) 上均匀分布，故还有 $y=-\frac{1}{\lambda}\ln\xi$, 其次，由此还可推知： $\gamma=-\frac{1}{\lambda}\ln(\xi_1\cdots\xi_n)$ 有参数为 n,λ 的 Gamma 分布，这是由于，此分布是几个指数分布的卷积 (见 298 页), 因而是 $\sum_{i=1}^{n}\frac{-1}{\lambda}\ln\xi_i=-\frac{1}{\lambda}\ln(\xi_1\cdots\xi_n)$ 的分布，这里 $\xi_1,\cdots,\xi_n$ 是独立的均匀分布的随机变数.

间 Δt 内观察一普阿松流，例如观察一放射性物质，如果在 Δt 中放射的粒子数是奇数就记为 1, 是偶数就记为 0, 这样连续不断地记录下去就得到 $\{\eta_i\}$, η_i 是第 i 个时间区间内观察的记录；实际上，由 §2.4 定理 2

$$P(\eta=0)=\sum_{k=0}^{\infty} e^{-\lambda\Delta t}\frac{(\lambda\Delta t)^{2k}}{(2k)!}=\frac{1+e^{-2\lambda\Delta t}}{2}.$$

如 $\lambda\Delta t$ 充分大，这概率接近 $\frac{1}{2}$. 另一种方法是利用前人已造好的随机数表，这是从 $0,1,2,\cdots,9$ 共十个数字中等可能地独立地取出的数字组成的表. 从这表中任取出一数，如是奇数，就记为 1, 是偶数就记为 0, 继续下去就得出一列 $\{\eta_i\}$.

以上方法有共同的缺点：一是我们决不可能观察无数多次以得到无数多个 η_i; 二是记录速度太慢，不能满足计算机的快速要求. 即使计算机上有随机数发生器的设备，也不能避免这些困难；因此便产生了研究伪随机数的问题.

(三) **伪随机数**. 所谓伪随机数，是指按照一定的计算方法而产生的数，它们具有类似于自均匀变数独立抽样值的性质. 这些数既然是依照决定性的算法而产生的，便不可能是真正的随机数；虽然如此，如果计算方法选得恰当，它们便近似于相互独立和近似于均匀分布，这就是说，它们能通过数理统计中的独立性检验和均匀分布检验. 由于这些原因，人们便称它们为伪随机数. 为了保证产生的速度，计算方法应该相当简便，以便于在计算机上迅速实现.

一般说来，伪随机数常借助于递推公式来产生：如已知前面 k 个伪随机数，下一个便是这 k 个的函数：

$$x_n=f(x_{n-1},x_{n-2},\cdots,x_{n-k}),$$

k 称为此递推式的级数.

产生的方法很多，我们先来叙述其中最常用的一种 - 余数法. 令

$$y_{n+1}\equiv\lambda y_n\,(\mathrm{mod}\ M),\quad y_0=a; \tag{4}$$

$$x_n=y_nM^{-1}, \tag{5}$$

其中 a,λ,M 为任意三个正常数， a 为奇数.

这就是说，令 y_0 等于 a; 若 y_n 已选定，以 M 除 λy_n 后的余数记为 y_{n+1}; y_n/M 就是所需的 x_n. 由构造方法可见： $0\leqslant y_n\leqslant M, 0\leqslant x_n\leqslant 1$. 因此，不同的 y_i(因之 x_i) 至多只有 M 个. 这表示序列 $\{x_n\}$ 是有周期 L 的， $L\leqslant M$, 每隔 L 个不同的 x_i 后循环一次. 既然如此， $\{x_n\}$ 就不可能是真正的随机序列. 不过如果 L 充分大，在同一个周期内的数却可能通过统计中的独立性与均匀性检验，这就完全依赖于参数 a,λ,M 的选择，有一些参数可以使它们通过而大多数则不能. 至于如何选择参数，目前主要依靠在计算机上作实验. 从一些已公开发表的报道得知下列参数组较为适用：

$$a=1,\lambda=5^{17},M=2^{42},(L=2^{40}\approx10^{12});$$
$$a=1,\lambda=5^{13},M=2^{36},(L=2^{34}\approx2\cdot10^{10});$$
$$a=1,\lambda=7,M=10^{10},(L=5\cdot10^{7}).$$

每选定一组参数因而获得一列 $\{x_n\}$, 我们必须对它进行统计检验，目的是看它是否具有较好的独立性 (即随机性) 和均匀分布性. 为了后一目的可利用 §5.5 的 χ^2 检验或 k 检验；至于检验独立性的方法可参看 [33] 第六章第五节，或徐钟济："随机性检验", 电子计算机动态， 1962 年，第 6,7,8 期. 我们这里再介绍一个同时检验这两种性质的方法，它的根据是下列事实：

设 ξ_1,ξ_2 是独立的随机变数，都在 (0,1) 中均匀分布，则和 $\xi=\xi_1+\xi_2$ 在 (0,2) 中有三角分布，它的密度是：

$$f(x)=\begin{cases} x, & \text{当}0<x<1, \\ 2-x, & \text{当}1\leqslant x<2. \end{cases} \tag{6}$$

实际上，令

$$g(x)=\begin{cases} 1, & \text{如}x\in(0,1), \\ 0, & \text{如}x\overline{\in}(0,1), \end{cases} \tag{7}$$

则 $g(x)$ 是 ξ_1(或 ξ_2) 的密度；由 §2.8(7)

$$f(x)=\int_{-\infty}^{\infty}g(x-z)g(z)dz=\int_0^1 g(x-z)dz=G(x)-G(x-1)$$
$$=\begin{cases} x, & \text{当}0<x<1, \\ 2-x, & \text{当}1\leqslant x<2. \end{cases}$$

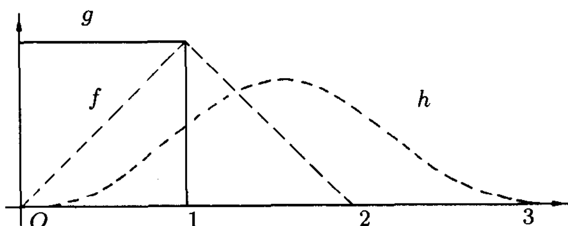


图 2

类似可证 (0,1) 中均匀分布的 n 重卷积密度是 $n-1$ 次多项式；特别，三重卷积密度 $h(x)$ 为

$$h(x)=\begin{cases} x^2/2, & \text{当}0<x<1, \\ [x^2-3(x-1)^2]/2, & \text{当}1\leqslant x<2, \\ [x^2-3(x-1)^2+3(x-2)^2]/2, & \text{当}2\leqslant x<3. \end{cases}$$

这就是所需要证明的事实.

现在设有一列由 (4)(5) 产生的数 $\{x_n\}$(为了使它有较好的性质，最好扔掉前面一些项而自某 x_k 取起，仍简记 x_k 为 x_1), 将它们两两相加得

$$\xi^{(1)}=x_1+x_2,\xi^{(2)}=x_3+x_4,\cdots,\xi^{(n)}=x_{2n-1}+x_{2n}.$$

然后用 χ^2(或 K) 检验，看 $\xi^{(1)},\xi^{(2)},\cdots,\xi^{(n)}$ 是否符合三角分布；如符合得好，那么可相信 $\{x_n\}$ 具有独立性和均匀分布性，否则就不能相信. 进一步，我们还可以三个相加或 k 个相加来进行检验，至于 k 个独立均匀变数的和的分布，见 [15] 卷二，第 27 页.

我们知道，统计检验不能十分肯定或否定某一假设；因此，只有让 $\{x_n\}$ 多通过几次检验才有较大的把握.

除 (4)(5) 而外，还有许多产生伪随机数的方法. 例如， (4) 式可换为

$$y_{n+1} \equiv \lambda y_n + b(\mathrm{mod}\, M), \quad y_0 = a, \tag{4$_1$}$$

这里多加了一个常数 $b > 0$.

又如取

$$y_{n+2} \equiv (y_{n+1} + y_n)\,(\mathrm{mod}\, M), \quad y_0 = 0,\ y_1 = 1;$$

$$x_n = y_n M^{-1}.$$

若取 $M = 2^{44}$, 则性质较好，此时周期约为 2.5×10^{13}.

再介绍一种方法：在 (0,1) 中任取二数 x_1, x_2, 将它们展为二进位数，设各有 $2n$ 项. 作乘积 x_1x_2, 它有 $4n$ 项，取中间 $2n$ 项 (即自第 $n+1$ 项取起至第 $3n$ 项止) 作为 x_3, 然后同样利用 x_2, x_3 作 x_4, 如此继续下去. 这里参数是 x_1, x_2.

不管用那种方法，都应该多作几次统计检验，以便尽可能最佳地选取参数，来保证伪随机数列具有较好的随机性质. (参看徐钟济："伪随机数之产生"及"随机性检验", 电子计算机动态， 1962, 第 5,6,7,8 期.)

§6.2 任意随机向量的模拟

(一) 反函数方法. 这一节的目的是要构造出具有事先给定的分布函数 $F(x)$ 的随机数，这种数称为 $F(x)-$ 随机数；如果 $F(x)$ 有密度 $f(x)$, 那么也称它们为 $f(x)-$ 随机数. 因此，上节所造的随机数是 (0,1) 均匀分布随机数，习惯上就简称为均匀随机数. 设随机变数 ξ 的分布函数是 $F(x)$, 所谓对 ξ 的模拟就是要构造若干个与 ξ 同分布的 $F(x)-$ 随机数；这种造法应该比较简单，以便能在计算机上实现. 以下会看到，一般地， $F(x)-$ 随机数都可自均匀随机数经过一些变换得到. 由此可见，均匀随机数是模拟的基石.

如果 $F(x)$ 是 n 元分布函数 $F(x_1, \cdots, x_n)$, 类似地可定义 $F(x_1, \cdots, x_n)-$ 随机向量；并可同样地考虑对 $\xi = (\xi_1, \cdots, \xi_n)$ 的模拟.

容易看出：如果 ξ 是 (0,1) 均匀分布随机数，那么

$$\eta = a + \xi(b - a),\ (b > a),$$

是 (a, b) 均匀分布随机数.

设已给分布函数 $F(x)$. 如果 $F(x)$ 连续，单调上升，则在 §6.1 中已提供了一个造 $F(x)-$ 随机数的方法：先取均匀随机数 x_i, 方程

$$F(y) = x_i \tag{1}$$

的解 $y_i = F^{-1}(x_i)$ 就是一个 $F(x)-$ 随机数. 如果 $x_1, x_2, \cdots, x_n$ 是独立的，那么， $y_1, y_2, \cdots, y_n$ 也独立.

这个方法在理论上虽然明确，但常常由于 $F(y)$ 的表达式较复杂而难于求解，因而不得不另觅途径. 下面我们叙述几个常用的方法，它们都不需要求 (1) 的解.

(二) **舍选法** (Neumann, J.). 这个方法的实质是从许多均匀随机数中选出一部分，使后者成为 $F(x)-$ 随机数.

引理 1 设 ξ,η 是两个独立的随机变数，ξ 在 (a,b) 中均匀分布而 η 在 (0,1) 中均匀分布；又设 $f(x)$ 是集中在 (a,b) 中的密度函数 (即是满足 $\int_a^b f(x)dx=1$ 的密度函数); 任取正常数 α, 使 $\alpha f(x)<1$ 对一切 $x\in(a,b)$ 成立，则有

$$P(\xi\leqslant d|\alpha f(\xi)\geqslant\eta)=\int_a^b f(x)dx,\ (a<d<b); \tag{2}$$

这就是说，在事件 $\alpha f(\xi)\geqslant\eta$ 下，ξ 的条件密度为 $f(x)$.

证 由假设，(ξ,η) 的密度函数为

$$g(x,y)=\frac{1}{b-a},\ \ x\in(a,b),\ y\in(0,1).$$

故

$$\begin{aligned}P(\xi\leqslant d|\alpha f(\xi)\geqslant\eta)&=\iint\limits_{\substack{\alpha f(x)\geqslant y\\ a<x\leqslant d}}\frac{dxdy}{b-a}\div\iint\limits_{\alpha f(x)\geqslant y}\frac{dxdy}{b-a}\\&=\int_a^d\int_0^{\alpha f(x)}dydx\div\int_a^b\int_0^{\alpha f(x)}dydx\\&=\int_a^d\alpha f(x)dx\div\int_a^b\alpha f(x)dx=\int_a^d f(x)dx.\ \square\end{aligned}$$

引理 1 启示我们造 $f(x)-$ 随机数的方法：先取一列 (a,b) 中均匀分布的随机数 $\xi_1,\xi_2,\cdots,\xi_n$, 再取一列均匀随机数 $\eta_1,\eta_2,\cdots,\eta_n$. 如果 $\alpha f(\xi_1)\geqslant\eta_1$, 就取 ξ_1 作为第一个 $f(x)-$ 随机数，如果此式不成立，就舍弃 ξ_1,η_1 而同样考查 ξ_2,η_2, 如此继续.

引理 1 中假定了 (a,b) 是有限区间. 如它不是有限的，我们总可选有限区间 (a,b), 使

$$\int_a^b f(x)dx\geqslant 1-\varepsilon, \tag{3}$$

其中 ε 是充分小的正数. 然后在 (a,b) 中运用上述方法，这时会出现小误差.

当 $f(x)$ 的变化不激烈时，此法较为有效；但如 $f(x)$ 在小区间内振幅很大，以致稍微改变 ξ_1 的值就会影响不等式 $\alpha f(\xi_1)\geqslant\eta_1$ 的成立时，这个方法就很不稳定. 其次，如 $f(x)$ 的表达式复杂，那么计算 $f(\xi_1)$ 也是繁重的工作.

(三) **离散逼近法**. 设密度 $f(x)$ 集中在 (a,b) 中，将 (a,b) 分成 n 等分，分点为 $a_0=a<a_1<a_2<\cdots<a_n=b$. 令

$$p_i=\int_{a_{i-1}}^{a_i}f(x)dx,\quad\left(\sum_{i=1}^n p_i=1\right).$$

将 $(0,1)$ 也分成 n 分，第 i 个区间的端点为 c_{i-1},c_i, 而且 $c_i-c_{i-1}=p_i$. 现在任取均匀随机数 η, 如 $c_{i-1}<\eta\leqslant c_i$, 则令

$$\xi=a_{i-1}+(a_i-a_{i-1})\frac{\eta-c_{i-1}}{c_i-c_{i-1}}.$$

如果 n 充分大，可以近似地认为 ξ 是 $f(x)-$ 随机数. 事实上

$$P(a_{i-1} < \xi \leqslant a_i) = P(c_{i-1} < \eta \leqslant c_i) = c_i - c_{i-1} = p_i = \int_{a_{i-1}}^{a_i} f(x)dx.$$

如果 $f(x)$ 不是集中在有限区间内，可以采用 (3) 中的截尾方法.

(四) 正态分布 $N(0,1)$ 随机变数的模拟. 除了可用离散逼近法外，还可以利用中心极限定理. 设 $\eta_1, \eta_2, \cdots, \eta_n$ 为 n 个独立的均匀变数，由于 $E\eta_i = \frac{1}{2}$, $D\eta_i = \frac{1}{12}$ 根据中心极限定理，当 k 相当大时， $\xi_k = \left(\sum\limits_{i=1}^{k} \eta_i - \frac{k}{2}\right) \Big/ \sqrt{\frac{k}{12}}$ 的分布渐近于 $N(0,1)$, 故可把 ξ_k 近似地取为 $N(0,1)-$ 随机数，通常可取 k 为 $8, 9, \cdots, 12$, 尤以 ξ_{12} 为最简便.

另一方法如下：取两个独立的均匀变数 η_1, η_2 作变换

$$\xi_1 = (-2\ln\eta_1)^{\frac{1}{2}} \cos 2\pi\eta_2,$$

$$\xi_2 = (-2\ln\eta_1)^{\frac{1}{2}} \sin 2\pi\eta_2,$$

则 ξ_1, ξ_2 是两个独立的 $N(0,1)$ 随机变数 [1]

(五) 离散型随机变数的模拟. 设有一列事件 $A_0, A_1, A_2, \cdots$ 满足完备性条件，即满足

$$A_iA_j = \phi\ (i \neq j); \quad P\left(\bigcup_{i=0}^{\infty} A_i\right) = 1.$$

我们来模拟这些事件. 令 $p_i = P(A_i)$, 则 $\sum\limits_{i=0}^{\infty} p_i = 1$. 又令 $l_k = \sum\limits_{i=0}^{k} p_i$ 于是 $\{l_k\}$ 将 $(0,1)$ 分成可列多个小区间. 现在任意取一均匀随机数 η, 如果 η 落在第 k 个小区间内，即如

$$l_{k-1} < \eta \leqslant l_k, \quad \text{就说出现事件} A_k. \tag{4}$$

这种模拟是合理的，因为

$$P(A_k) = p(l_{k-1} < \eta \leqslant l_k) = l_k - l_{k-1} = p_k.$$

由此可见，对一列完备事件的试验，化为对均匀变数的试验. 我们称 (4) 为事件模拟法则，它有较多的应用.

现在来模拟离散型随机变数 ξ, 它有密度矩阵为 $\begin{pmatrix} x_0 & x_1 & x_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}$. 显然事件 $A_i = (\xi = x_i), i = 0, 1, 2, \cdots$ 满足完备性条件，于是可运用上述法则来模拟 ξ, 即如 $\sum\limits_{i=0}^{k-1} p_i < \eta \leqslant \sum\limits_{i=0}^{k} p_i$, 就取对应于 ξ 的随机数为 x_k. (令 $\sum\limits_{i=0}^{-1} p_i = 0$.) [2]

---

[1] 证：设 ξ_1, ξ_2 独立，同 $N(0,1)$ 分布，则 $Q = (\xi_1, \xi_2)$ 的密度为 $p_Q(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) = \frac{1}{2\pi}\exp\left(-\frac{r^2}{2}\right)$, 作变换 $x = r\cos\varphi, y = r\sin\varphi$. 变换行列式等于 r, 故 $p_Q(x,y) = p_Q(r,\varphi) = r\exp\left(-\frac{r^2}{2}\right)\cdot\frac{1}{2\pi}$. 这表示 Q 的极坐标 R, Φ 独立，分别有密度为 $r\exp\left(-\frac{r^2}{2}\right)$ 及 $\frac{1}{2\pi}$. 取二独立均匀变数 η_1, η_2 对 R 及 Φ 用反函数法，由 $\int_0^R r\exp\left(-\frac{r^2}{2}\right)dr = 1 - \eta_1$ (后者也是均匀变数) 及 $\int_0^{\Phi} \frac{d\varphi}{2\pi} = \eta_2$, 解得 $R = \sqrt{-2\ln\eta_1}, \Phi = 2\pi\eta_2$. 于是 $\xi_1 = R\cos\Phi = (-2\ln\eta_1)^{\frac{1}{2}}\cos 2\pi\eta_2$; 类似讨论 ξ_2.

[2] 如 ξ 有二项分布 $B(n,p)$, 还可用下法模拟：取 n 个均匀随机数 $\eta_1, \cdots, \eta_n$, 若其中恰有 k 个不大于 p, 就令 $\xi = k$, 易见 $P(\xi = k) = \binom{n}{k} p^k(1-p)^{n-k}$.

(六) **随机向量的模拟**. 现在来模拟 n 维随机向量 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)$, 它的概率密度为 $f(x_1,x_2,\cdots,x_n)$. 我们先回忆 §1.4 定理 2 中的乘法公式：设 $A_1,\cdots,A_n$ 为 n 个事件, $P(A_1A_2\cdots A_n)>0$, 则有

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}). \tag{5}$$

与此类似，对多维密度函数也有公式

$$f(x_1,x_2,\cdots,x_n)=f_1(x_1)f(x_2|x_1)f(x_3|x_1,x_2)\cdots f(x_n|x_1,x_2,\cdots,x_{n-1}); \tag{6}$$

其中 $f_1(x_1)$ 是 ξ 的边沿分布密度，而 $f(x_k|\ x_1,x_2,\cdots,x_{k-1})$ 是在 $\xi_1=x_1,\xi_2=x_2,\cdots,\xi_{k-1}=x_{k-1}$ 的条件下， ξ_k 的条件分布密度，它们都是一维密度函数. 为了证 (6), 令

$$\left.\begin{aligned}
&f_{n-1}(x_1,\cdots,x_{n-1})=\int f(x_1,\cdots,x_n)dx_n,\\
&f_{n-2}(x_1,\cdots,x_{n-2})=\int f_{n-1}(x_1,\cdots,x_{n-1})dx_{n-1},\\
&f_{n-3}(x_1,\cdots,x_{n-3})=\int f_{n-2}(x_1,\cdots,x_{n-2})dx_{n-2},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&f_1(x_1)=\int f_2(x_1,x_2)dx_2,
\end{aligned}\right\} \tag{7}$$

其中积分域是 $(-\infty,\infty)$. 由 §2.6 (三) 知 $f_k(x_1,\cdots,x_k)$ 是 $(\xi_1,\cdots,\xi_k)$ 的 k 维联合分布密度. 由 §2.7(二),

$$f(x_k|x_1,\cdots,x_{k-1})=\frac{f_k(x_1,\cdots,x_k)}{f_{k-1}(x_1,\cdots,x_{k-1})}, \tag{8}$$

以 (8) 代入 (6) 式右方，即得证 (6) 式成立.

(6) 式启示我们模拟 ξ 的方法：利用 (一) 至 (四) 中的方法，先造出 $f_1(x_1)$- 随机数 y_1; 以 y_1 代入 $f(x_2|x_1)$ 中的 x_1, 得到一个确定的一维密度 $f(x_2|y_1)$, 再造出 $f(x_2|y_1)$- 随机数 y_2; 如 $y_1,\cdots,y_{k-1}$) 已确定，再造出 $f(x_k|y_1,\cdots,y_{k-1})$- 随机数 y_k. 所得的 n 维向量 $(y_1,\cdots,y_n)$ 就是 $(\xi_1,\cdots,\xi_n)$ 的模拟.

对离散型随机向量的模拟可类似作出.

这个方法的实质是化多维的模拟为一维模拟，其优点是它的普遍性，缺点是要计算 (7) 中 $n-1$ 个积分，这对复杂的函数难以实现.

如果容许一定的误差，我们也可以仿用 (三) 中的离散逼近法，这时可以避免积分的计算，只要 $f(x_1,\cdots,x_n)$ 的变化比较缓慢，因为 ξ 落在 n 维空间一小区域 ΔV 内的概率可以用 $f(x_1^0,\cdots,x_n^0)|\Delta V|$ 来代替，这里 $(x_1^0,\cdots,x_n^0)\in\Delta V$, 而 $|\Delta V|$ 是 ΔV 体积.

(七) **n 维正态分布随机向量的模拟**. 对这种随机向量的模拟有下列简便方法. 设 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)$ 的密度为 (参看 §2.6 例 7)

$$f(x_1,x_2,\cdots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}e^{-\frac{1}{2}XB^{-1}X'},$$

其中 $X=(x_1,x_2,\cdots,x_n)$ 为 n 维向量， $B=(b_{jk})$ 为 n 级正定对称矩阵. 由 §2.12 定理 2 知

$$E\xi_j=0,\quad E\xi_j\xi_k=b_{jk}. \tag{9}$$

现在任取 n 个相互独立的一维 $N(0,1)$ 正态分布随机变数 $\zeta_1,\zeta_2,\cdots,\zeta_n$, 并作三角形变换

$$\begin{aligned}&\xi_1'=a_{11}\zeta_1,\\&\xi_2'=a_{21}\zeta_1+a_{22}\zeta_2,\\&\cdots\cdots\\&\xi_n'=a_{n1}\zeta_1+a_{n2}\zeta_2+\cdots+a_{nn}\zeta_n,\end{aligned} \tag{10}$$

由于 $\xi'=(\xi_1',\xi_2',\cdots,\xi_n')$ 由正态随机向量 $(\zeta_1,\zeta_2,\cdots,\zeta_n)$ 经线性变换得来，故 ξ' 也是正态分布向量. 如果我们能适当选取变换系数 $a_{11},a_{21},a_{22},\cdots,a_{nn}$ 使 ξ' 的二级矩矩阵也是 B, 亦即使

$$E\xi_j'=0,\quad E\xi_j'\xi_k'=b_{jk}, \tag{11}$$

则因正态分布由前二级矩唯一决定，故由 (9)(11) 知 ξ 与 ξ' 有相同的分布，因而由 (10) 造出的 ξ' 是 ξ 的模拟，而 $\zeta,\cdots,\zeta_n$ 的产生方法是我们已知的.

以下求 a_{ij}. 由 $\{\zeta_i\}$ 的独立性及 $E\zeta_j=0$,

$$E\zeta_j\zeta_k=\begin{cases}0, & j\neq k,\\ 1, & j=k.\end{cases}$$

由 (11) 及 (10) 得

$$\begin{aligned}b_{jk}&=E\left(\sum_{l=1}^{j}a_{jl}\zeta_l\right)\left(\sum_{m=1}^{k}a_{km}\zeta_m\right)\\&=\sum_{l=1}^{s}a_{jl}a_{kl},\ (s=\min(j,k)).\end{aligned} \tag{12}$$

特别

$$\begin{aligned}&b_{11}=a_{11}^2,\ \ 故\ a_{11}=\sqrt{b_{11}};\\&b_{12}=a_{11}a_{21},\ \ 故\ a_{21}=\frac{b_{12}}{\sqrt{b_{11}}},\\&b_{22}=a_{21}^2+a_{22}^2,\ \ 故\ a_{22}=\sqrt{\frac{b_{11}b_{22}-b_{12}^2}{b_{11}}};\end{aligned} \tag{13}$$

一般，如已求出 $a_{11},\cdots,a_{k1},a_{k2},\cdots,a_{kk}$, 则可由 (12) 求出 $a_{k+11},a_{k+12},\cdots,a_{k+1k+1}$, 于是 (10) 中系数完全决定.

现在来考虑一个特例：设

$$b_{jk}=\sigma^2e^{-d|j-k|},\ (\sigma>0,d>0\ 为常数). \tag{14}$$

为简便计先设 $\sigma=1$. 此时由 (13) 得

$$a_{j1}=e^{-(j-1)d},\ a_{jl}=e^{-(j-l)d}\cdot\sqrt{1-e^{-2d}},\ (2\leqslant l\leqslant j).$$

代入 (10), 得

$$\begin{aligned}
\xi_1' &= \zeta_1,\\
\xi_2' &= e^{-d}\zeta_1 + \sqrt{1-e^{-2d}}\cdot\zeta_2 = e^{-d}\xi_1' + \sqrt{1-e^{-2d}}\cdot\zeta_2,\\
\xi_3' &= e^{-d}(e^{-d}\zeta_1 + \sqrt{1-e^{-2d}}\cdot\zeta_2) + \sqrt{1-e^{-2d}}\cdot\zeta_3\\
&= e^{-d}\xi_2' + \sqrt{1-e^{-2d}}\cdot\zeta_3.
\end{aligned}$$

一般有

$$\xi_j' = e^{-d}\xi_{j-1}' + \sqrt{1-e^{-2d}}\cdot\zeta_j.$$

以上假定了 $\sigma=1$. 对一般的 $\sigma>0$, 易见

$$\left.\begin{aligned}
\xi_1' &= \sigma\zeta_1,\\
\xi_j' &= e^{-d}\xi_{j-1}' + \sigma\sqrt{1-e^{-2d}}\cdot\zeta_j,\ (1<j\leqslant n).
\end{aligned}\right\} \tag{15}$$

这是一个递推公式，对电子计算机特别合适. 当已算得 ξ_{j-1}' 后，只要再抽取一个 $N(0,1)$–随机数 ζ_j, 作 ξ_{j-1}' 与 ζ_j 的线性组合就得 ξ_j'.

可惜，对一般的 (b_{jk}), 未必能经常得到类似的递推式. 进一步的研究请参看论文:

1) В. Г. Срагович: Моделирование некоторых классов случайных процессов, Журнал вычислительной математики и математической Физики, 1963, 3:3, 586-593.

2) J. N. Franklin: Numerical simulation of stationary and nonstationary Gaussian random processes, SIAM Review, vol.7, N0.1, 1965, 68-80.

§6.3 随机过程的模拟

(一) 马尔科夫链的模拟. 设已给随机矩阵

$$P=(p_{ij}),\qquad (i,j=0,1,2,\cdots).$$

我们来模拟一个马尔科夫链，它的状态空间是 $E=(0,1,2,\cdots)$, 一步转移概率矩阵是 P, 而且具有开始分布为 $q=(q_0,q_1,q_2,\cdots)$. 为此，采用 §6.2 中的事件模拟法则，如下定义一列随机变数 $\xi_0,\xi_1,\xi_2,\cdots$:

1) 任取一均匀随机数 η_0, 如果 $\sum\limits_{i=0}^{k_0-1} q_i < \eta_0 \leqslant \sum\limits_{i=0}^{k_0} q_i$, 就定义 $\xi_0=k_0$;

2) 再取一均匀随机数 η_1, 如果 $\sum\limits_{i=0}^{k_1-1} p_{k_0 i} < \eta_1 \leqslant \sum\limits_{i=0}^{k_1} p_{k_0 i}$, 就定义 $\xi_1=k_1$;

3) 一般，设 $\xi_j=k_j$ 已确定，再取一均匀随机数 η_{j+1}, 如果 $\sum\limits_{i=0}^{k_{j+1}-1} p_{k_j i} < \eta_{j+1} \leqslant \sum\limits_{i=0}^{k_{j+1}} p_{k_j i}$, 就定义 $\xi_{j+1}=k_{j+1}$.

这样造出的 $\{\xi_j\}$ 就是所需的马尔科夫链. 这个方法还可以用来模拟非齐次的、具可列多个状态的马尔科夫链.

下面考虑时间参数连续的情况，一类重要的马氏过程的运动法则由两种随机因素决定：一是状态间的转移，它象离散参数的马尔科夫链一样，由一个随机矩阵 $P=(p_{ij})$,

$(i,j=0,1,2,\cdots)$ 所描述，不过这时总有 $p_{ii}=0$(不能立刻回到自己); 二是停留时间的分布，它总是指数型的，即

$$P_i(\tau \leqslant t) = 1 - e^{-\lambda_i t}, \ (t \geqslant 0). \tag{1}$$

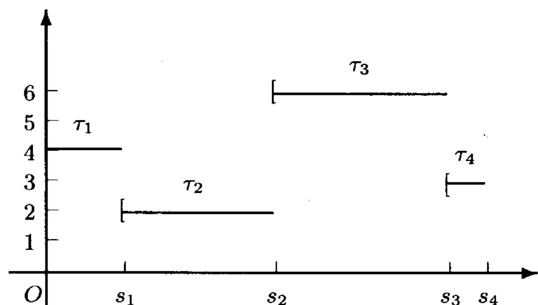


图 1 马尔科夫链的一个样本函数图，记号 ⊢ 表右连续

左方记号的意义是：在开始位于状态 i 的条件下，停留时间 τ 不超过 t 的条件概率.

这种过程可如下模拟：

1° 如上 1), 定义

$$\xi_0 = k_0;$$

任取一个具有参数为 λ_{k_0} 的指数分布随机数 τ_1(参看 §6.1 例 1);

2° 如上 2), 定义 $\xi_1 = k_1$; 任取一个参数为 λ_{k_1} 的指数分布随机数 τ_2;

3° 如上 3), 定义 $\xi_{j+1} = k_{j+1}$, 任取一参数为 $\lambda_{k_{j+1}}$ 的指数分布随机数 τ_{j+2};

4° 如非负数 $t \in \left[\sum\limits_{i=0}^{l} \tau_i, \sum\limits_{i=0}^{l+1} \tau_i\right]$,(补令 $\tau_0 = 0$) 定义 $X_t = k_l$.

这样造出的 $\{X_t, t \geqslant 0\}$ 就是所需的过程. 上述造法的直观意义是：质点自 k_0 出发，停留一段时间 τ_1, 转移到 k_1, 又停留一段时间 τ_2, 转移到 $k_2, \cdots$. 特别，如果 P 满足 $p_{ii+1} = 1$, $p_{ij} = 0, (j \neq i+1, i = 0,1,2,\cdots)$, 则所得 $\{X(t), t \geqslant 0\}$ 是普阿松过程 (参看 §2.4 及 §4.2 例 5).

(二) 弱平稳过程的模拟. 我们来模拟一个具有已给相关函数为 $B(\tau), (\tau \in R_1)$ 的弱平稳过程 $\{X(t), t \in R_1\}$, 而且 $EX(t) \equiv 0$. 自然，在计算机上只能造出过程在有限多个等距离的点 $t_1, \cdots, t_n$ 上的值， $t_{i+1} - t_i = k$(常数). 把 $X(t_i)$ 表为

$$\left.\begin{aligned}
X(t_1) &= c_1\eta_1 + c_2\eta_2 + \cdots + c_n\eta_n, \\
X(t_2) &= c_1\eta_2 + c_2\eta_3 + \cdots + c_n\eta_{n+1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
X(t_k) &= c_1\eta_k + c_2\eta_{k+1} + \cdots + c_n\eta_{n+k-1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
X(t_n) &= c_1\eta_n + c_2\eta_{n+1} + \cdots + c_n\eta_{2n-1},
\end{aligned}\right\} \tag{2}$$

其中 $\{\eta_i\}$ 为任意满足 $E\eta_i = 0$, 方差 $D\eta_i = B(0), E\eta_i\eta_j = 0, (i \neq j)$ 的随机变数，它们可按 §6.2 中方法产生. $\{c_i\}$ 为待定常数. 因

$$EX(t_k)X(t_1) = E\left(\sum_{i=1}^{n} c_i\eta_{i+k-1}\right)\left(\sum_{i=1}^{n} c_i\eta_i\right) = B(0)\cdot\sum_{i=1}^{n-k+1} c_ic_{i+k-1},$$

而且要求 $B(\tau) = EX(t+\tau)X(t)$, 故应从下列 n 个方程

$$B(t_k - t_1) = B(0)\cdot(c_1c_k + c_2c_{k+1} + \cdots + c_{n-k+1}c_n), \ (k = 1,\cdots,n) \tag{3}$$

解出 $c_1, c_2, \cdots, c_n$ 代入 (2) 即得 $X(t_1), \cdots, X(t_n)$. 易见 $\{X(t_i)\}$ 是弱平稳的；实际上，我们有

$$\begin{aligned} EX(t_{k+i})X(t_i) &= E(c_1\eta_{k+i} + c_2\eta_{k+i+1} + \cdots + c_n\eta_{n+k+i-1})(c_1\eta_i + c_2\eta_{i+1} + \cdots + c_n\eta_{n+i-1}) \\ &= B(0)[c_1c_{k+1} + c_2c_{k+2} + \cdots + c_{n-k}c_n] = B(t_{k+1} - t_1) = B(kh) = B(t_{k+i} - t_i), \end{aligned}$$

后者只依赖于 t_i, t_{k+i} 间的距离 $t_{k+i} - t_i$.

这里所需的 $\{\eta_i\}$ 可取为独立的在 $(-h, h)$ 中均匀分布的随机数，其方差为 $\frac{h^2}{3} = B(0)$, $(\therefore h = \sqrt{3 \cdot B(0)})$ 或取为独立的 $N(0, \sqrt{B(0)})$ 正态随机数.

本方法的缺点在于解 (3) 时计算量大；注意 (3) 是二次方程组，一般只能求出近似解.

(三) **平稳正态马尔科夫过程的模拟**. 现在来模拟在实际中很有用处而且重要的过程 $\{X(t), t \geqslant 0\}$, 它是实值弱平稳相关函数为

$$B(\tau) = \sigma^2 e^{-\beta|\tau|}, \ (\sigma > 0, \beta > 0 \text{ 常数}) \tag{4}$$

的正态过程. 在 §4.5 例 4 中已谈到, 所谓 $\{X(t)\}$ 是正态过程, 是指对任意 n 个 $t_1 < t_2 < \cdots < t_n$, $X(t_1), X(t_2), \cdots, X(t_n)$ 的联合分布是 n 维正态分布. 为简单起见，我们设. $EX(t) \equiv 0$. 于是这个 n 维正态分布只由协方差矩阵 $B = (b_{jk})$ 决定. 由于 (4)

$$b_{jk} = EX(t_j)X(t_k) = B(t_j - t_k) = \sigma^2 e^{-\beta|t_j - t_k|}. \tag{5}$$

根据 §4.5 例 3', 我们知道 (4) 所对应的谱密度为

$$f(x) = \frac{\sigma^2}{\pi} \cdot \frac{\beta}{x^2 + \beta^2}. \tag{6}$$

由 §4.5 例 4, 我们还知道这过程是平稳过程. 其实它还是马尔科夫过程，不过我们不在这里证明这一结论 (参看 [35]§9.3).

由 (5) 看出，当 $\beta \to \infty$ 时， $EX(t_j)X(t_k) = 0$, 意即 $X(t_j)$ 与 $X(t_k)$ 不相关，既然过程是正态的， $X(t_j)$ 与 $X(t_k)$ 独立；而且由 (6), 谱密度近似于一个常数 $\frac{\sigma^2}{\pi\beta}$. 在工程书中称相互独立而且谱密度等于正常数的过程为白噪声，常被用来模拟噪声或干扰. 从数学观点看来，白噪声不是弱平稳过程，因为后者的谱密度可积分 (参看 §4.5(24)), 而白噪声的谱密度既是一个正常数，其积分等于 ∞.

模拟 $\{X(t), t \geqslant 0\}$, 实际上只能模拟有穷或可列多个随机变数 $X_n = X(n \cdot \Delta t), n = 1, 2, \cdots$, 步长 $\Delta t > 0$ 为一常数. 为此，采取下列步骤：

1) 产生在 (0,1) 中均匀分布的伪随机数 $\{u_n\}$:

$$y_{n+1} \equiv \lambda y_n, \ (\text{mod} \cdot m), \ \ y_0 = a; \ \ u_n = y_n \cdot M^{-1} \tag{7}$$

(见 §6.1(三)).

2) 利用 $\{u_n\}$ 造一列 $N(0,1)$ 正态分布随机数 $\{w_n\}$, 这可用 §6.2(四) 中的方法来造. 例如，一种方法是令

$$\omega_{2n-1} = (-2\ln u_{2n-1})^{\frac{1}{2}} \cos 2\pi u_{2n},$$

$$\omega_{2n} = (-2\ln u_{2n-1})^{\frac{1}{2}} \sin 2\pi u_{2n},$$

于是一对 u_n 产生一对 ω_n.

3) 根据 §6.2(15), 由 $\{\omega_n\}$ 产生 $\{X_n\}$:

$$\left.\begin{aligned}&X_1=\sigma\omega_1,\\&X_n=e^{-\beta\Delta t}X_{n-1}+\sigma\sqrt{1-e^{-2\beta\Delta t}}\cdot\omega_n.\end{aligned}\right\}\tag{8}$$

由 §6.2 中推导此式的根据与程序，可知这样的 $\{X_n\}$ 满足上述要求.

问题在于：作为出发点的 u_n, 并不是真正的 (0,1) 均匀分布随机数，而是按决定性公式 (7) 所产生的伪随机数，它们是否具有较好的独立性与均匀分布性，还需要进行统计检验；对 $\{\omega_n\}$ 也有独立性与正态性的检验问题；至于 $\{X_n\}$, 则需要检验它是弱平稳、正态而且相关函数为

$$EX_{n+\tau}X_n=\sigma^2e^{-\beta|\tau|\Delta t}\tag{9}$$

的序列.

对 $\{u_n\}$ 的检验法已在 §6.1 中提过. 对 $\{\omega_n\}$, 可用 §5.5 中 χ^2(或 K) 检验法来研究它的正态性，至于检验独立性，我们再补充一个方法.

记 $y_i=\omega_{2i-1}$, $z_i=\omega_{2i}, i=1,2,\cdots,n$, 令

$$\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,\quad \overline{z}=\frac{1}{n}\sum_{i=1}^{n}z_i.$$

子样的相关系数为

$$r=\frac{\frac{1}{n}\sum(y_i-\overline{y})(z_i-\overline{z})}{\sqrt{\frac{1}{n}\sum(y_i-\overline{y})^2}\sqrt{\frac{1}{n}\sum(z_i-\overline{z})^2}}=\frac{\frac{1}{n}\sum y_iz_i-\overline{y}\,\overline{z}}{\sqrt{\frac{1}{n}\sum y_i^2-\overline{y}^2}\sqrt{\frac{1}{n}\sum z_i^2-\overline{z}^2}},$$

其中 $\sum$ 代表 $\sum\limits_{i=1}^{n}$. 如果 $\{\omega_i\}$ 独立，则有

$$P(\sqrt{(1.96)^2+n-2}\cdot|r|>1.96)\approx 5\%,\ (n>120).$$

如以 2.04 换 1.96, 则只需 $n>30$. 因此，如 $\{\omega_i\}$ 使上式左方括号中不等式成立，就应否定"独立性"假设. 我们不可能给出检验法的证明，因为每种检验法都以某定理为后盾，为证明该定理则需要大量的篇幅，并且会使我们远离正题. 我们只指明定理的出处，从那里可以找到证明. 上述检验法见 [3],§29.7.

对 $\{X_n\}$ 的检验，可以举出很多，这里只介绍四种：如 $\{X_n\}$ 是弱平稳、正态而且满足 (9), 则下诸结果正确：

1)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_iX_{i+\tau}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-\tau}X_iX_{i+\tau}=\sigma^2e^{-\beta|\tau|\Delta t};$$

$$\lim_{n\to\infty}\frac{\frac{1}{n}\sum\limits_{i=1}^{n}X_iX_{i+\tau}-\overline{X}_0\overline{X}_\tau}{s_0\cdot s_\tau}=\sigma^2e^{-\beta|\tau|\Delta t};\ \text{(a.s.)}$$

这里

$$\overline{X}_\tau = \frac{1}{n}\sum_{i=1}^{n} X_{i+\tau};$$

$$s_\tau^2 = \frac{1}{n}\sum_{i=1}^{n} X_{i+\tau}^2 - \overline{X}_\tau^2.$$

(见 [16] 第十章 §7). 此外, 作为 $e^{-\beta|\tau|\Delta t}$ 的最大似然估值可取 $\sum\limits_{i=1}^{n} X_iX_{i+1}\Big/\sum\limits_{i=1}^{n} X_i^2$(见 [20]§8.3).

2)

$$P\left(\frac{\left|\sum\limits_{i=1}^{n}(X_i - e^{-\beta\Delta t}X_{i-1})X_{i-1}\right|}{\sigma^2\cdot\sqrt{(1-e^{-2\beta\Delta t})n}} \leqslant 1.96\right) \approx 95\%.$$

(见 [20]§8.3).

3) 令 $z_n = \frac{1}{\sigma}\max(X_1, X_2, \cdots, X_n)$, 则有 [1]

$$\lim_{n\to\infty} P\left(\frac{z_n - b_n}{a_n} \leqslant x\right) = e^{-e^{-x}};$$

故

$$P(-1.25 < \frac{z_n - b_n}{a_n} < 3.5) \approx 94\%$$

(令 $F(x) = e^{-e^{-x}}$, 则 $F(3.51) \approx 97\%, F(-1.26) \approx 3\%$). 这里 $a_n = (2\ln n)^{-\frac{1}{2}}$; $b_n = (2\ln n)^{\frac{1}{2}} - \frac{1}{2}(2\ln n)^{-\frac{1}{2}}(\ln\ln n + \ln 4\pi)$. (见 Annals of Mathematical Statistics,1964, 第 508 页及 [3] §28.6).

4)

$$\lim_{n\to\infty} P\left(\frac{\sum\limits_{i=1}^{n+1} X_i}{\left[E\left(\sum\limits_{i=1}^{n+1} X_i\right)^2\right]^{\frac{1}{2}}} \leqslant y\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{x^2}{2}}dx,$$

其中

$$\begin{aligned}
E\left(\sum_{i=1}^{n+1} X_i\right)^2 &= E\left(\sum_{i=1}^{n+1} X_i^2\right) + 2E\left(\sum_{i=1}^{n}\sum_{j>i}^{n+1} X_iX_j\right)\\
&= (n+1)\sigma^2 + 2\sigma^2\left[\sum_{k=1}^{n} e^{-k\beta\Delta t} + \sum_{k=1}^{n-1} e^{-k\beta\Delta t} + \cdots + e^{-\beta\Delta t}\right]\\
&= (n+1)\sigma^2 + \frac{2\sigma^2 e^{-\beta\Delta t}}{1-e^{-\beta\Delta t}}\left[n - \frac{e^{-\beta\Delta t}(1-e^{-n\beta\Delta t})}{1-e^{-\beta\Delta t}}\right].
\end{aligned}$$

(见 Теория вероятностей и её применения, 1959, 4:1, 186-207 页; [38] 391 页及 459 页).

[1]称 $F(x) = \exp(-\alpha e^{-\beta x})$ 为重指数分布函数, 参数 $\alpha > 0, \beta > 0$. 它是极大值的一种极限分布, 在洪水、地震预报及可靠性问题中有用, 见 [8,41]. 不难证明: 如 $\eta, \xi_1, \xi_2, \cdots$ 独立, η 有参数为 α 的普阿松分布, 各 ξ_i 有相同的指数分布, 参数为 β, 则 $\sum\limits_{i=1}^{\eta} \xi_i$ 有重指数分布. 证明可仿 §2.9 例 5.

(四) 维纳过程 (布朗运动) 的模拟. 这类过程 $\{X(t), t \geqslant 0\}$ 已在 §4.2 例 6 中介绍过. 设取样步长为 $\Delta t > 0$, 令 $X_n = X(n \cdot \Delta t)$. 受 (8) 式的启发, 可以想到, $\{X_n\}$ 可由下列递推式产生:

$$\left.\begin{aligned} &X_1 = \sigma \cdot \sqrt{\Delta t} \cdot \omega_1, \\ &X_n = X_{n-1} + \sigma \cdot \sqrt{\Delta t} \cdot \omega_n, (\sigma > 0). \end{aligned}\right\} \tag{10}$$

(设 $X_0 = 0$), 其中 $\{\omega_n\}$ 为相互独立的 $N(0,1)$ 随机数. 实际上, 由 (10) 中第二式

$$\begin{aligned} X_n &= X_{n-2} + \sigma \cdot \sqrt{\Delta t} \cdot \omega_{n-1} + \sigma \cdot \sqrt{\Delta t} \cdot \omega_n = X_{n-2} + \sigma \cdot \sqrt{\Delta t}(\omega_{n-1} + \omega_n) \\ &= \cdots = X_{n-k} + \sigma \cdot \sqrt{\Delta t}(\omega_{n-k+1} + \omega_{n-k+2} + \cdots + \omega_n), \end{aligned} \tag{11}$$

故增量 $X_n - X_{n-k}$ 只依赖于对应于 $(n-k, n]$ 的 k 个变量 $(\omega_{n-k+1}, \cdots, \omega_n), (k < n)$. 既然由假定 $\{\omega_n\}$ 相互独立, 可见对 $n_1 < n_2 < \cdots < n_l$, 诸增量

$$X_{n_2} - X_{n_1}, X_{n_3} - X_{n_2}, \cdots, X_{n_l} - X_{n_{l-1}}$$

相互独立. 其次, 由 (11) 及 ω_i 有 $N(0,1)$ 分布的假定, 知 $X_n - X_{n-k}$ 有 $N(0, \sigma \cdot \sqrt{k\Delta t})$ 分布. 这便证明了 $\{X_n\}$ 是维纳过程在点列 $n\Delta t$ 上的抽样.

*第七章　概率论在计算方法中的一些应用

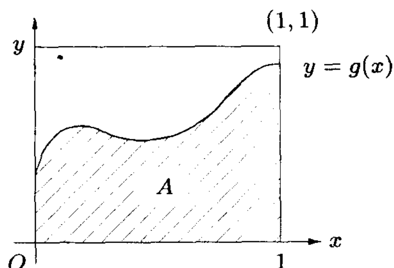
§7.1　定积分的计算

(一) 掷点算法. 在本章中，我们对概率计算方法作些初步的介绍，以作为基本理论的一种应用. 由于电子计算机的出现，近年来发展了用概率模型来作近似计算的方法，通常称为蒙特卡罗 (Monte-Garlo) 方法. 这种方法的基本思想是：为了要计算某些量，先造出概率模型 (如随机向量、随机过程等), 使后者的若干数字特征 (例如数学期望), 恰好重合于所需计算的量；而这些数字特征，又可以通过实验，用统计方法求出它们的估值；于是人们便把这些估值作为要求的量的近似值.

这里自然地发生三个问题：第一，如何造概率模型？这需要具体问题具体解决. 这些模型应该充分简单，以便能在手头上或计算机上实现. 第六章的内容，便是解决这个问题的基础. 第二，既然是近似计算，便有误差的估计问题. 由于我们是以统计估值作为近似值，自然希望估值的方差越小越好. 第三是计算的速度问题，运算的次数和时间越少越好.

设 $g(x)$ 是有限区间 $[a,b]$ 上的连续函数，需要计算定积分

$$G=\int_a^b g(x)dx \tag{1}$$



的值. 在 §3.3 例 1 中已提出了一种算法，它的直观想法是：经过简单的线性变换，不妨设 $[a,b]=[0,1], 0\leqslant g(x)\leqslant 1$. 于是 G 等于图 1 中的面积 A. 现在向矩形 $0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1$ 中均匀分布地掷点 $P_i=(\xi_i,\eta_i), i=1,2,\cdots,n$, 落于 A 中的点数设为 $k(=k(n))$, 则由于每次成功 (即落于 A 中) 的概率为 G, 故由强大数定理，有

$$\lim_{n\to\infty}\frac{k}{n}=G.$$

所谓均匀分布地掷点应理解为 $\xi_1,\eta_1,\xi_2,\eta_2,\cdots$ 是独立的均匀随机数，而第 i 次获得成功则等价于

$$g(\xi_i)\geqslant\eta_i. \tag{2}$$

既然 k 是 n 次贝努里试验 (即掷点试验) 中成功的总次数，故 k 有二项分布 $B(n,G)$, 方差为 $D(k)=nG(1-G)$, 因此

$$D\left(\frac{k}{n}\right)=\frac{1}{n^2}D(k)=\frac{G(1-G)}{n}. \tag{3}$$

而数学期望为

$$E\left(\frac{k}{n}\right)=\frac{1}{n}E(k)=G.$$

关于精确度和运算次数，我们提出下列问题：应取多大的 n, 才能使 $\left|\frac{k}{n}-G\right|<\varepsilon$ 的概率不小于 A? 这里 ε 为充分小而 A 为接近于 1 的正数. 根据中心极限定理 (见 §3.4 系 1), $\frac{k-nG}{\sqrt{nG(1-G)}}$ 渐近地有 $N(0,1)$ 正态分布，因此

$$P\Big(\Big|\frac{k}{n}-G\Big|<\varepsilon\Big)=P\Big(\Big|\frac{k-nG}{\sqrt{nG(1-G)}}\Big|<\varepsilon\cdot\sqrt{\frac{n}{G(1-G)}}\Big)$$
$$\approx\Phi\Big(\varepsilon\cdot\sqrt{\frac{n}{G(1-G)}}\Big)-\Phi\Big(-\varepsilon\cdot\sqrt{\frac{n}{G(1-G)}}\Big)\geqslant A,\tag{4}$$

其中 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy$. $\Phi(x)$ 之值有正态分布表可查. 把 ε, G, A 看成已知数，即可解出满足 (4) 的最小的 n 来；实际上，查表，对任何 $A, 0<A<1$, 可找到 $t_A>0$, 使

$$\Phi(t_A)-\Phi(-t_A)\approx A,$$

例如：当 $A=99.7\%, 95.6\%, 68.3\%$ 时， t_A 分别为 3,2,1(参看 §2.5 图 1), 于是 (4) 之解化为解

$$\varepsilon\cdot\sqrt{\frac{n}{G(1-G)}}=t_A.$$

由此得

$$n\approx\frac{t_A^2G(1-G)}{\varepsilon^2}.\tag{5}$$

特别，当 $A=99.7\%$ 时， $n\approx\frac{9G(1-G)}{\varepsilon^2}$.

但这里出现了矛盾， G 是待求的数，是未知的，怎样由此算出 n 呢？我们可以先试算若干次，以求出 G 的一个粗略估值 $\widehat{G}$, 然后以此 $\widehat{G}$ 代入上式中的 G 来估计 n.

例 计算积分

$$G=\int_0^1\frac{e^x}{4}dx.\tag{6}$$

由直接积分可得

$$G=\frac{1}{4}(e-1)=0.42957\cdots\approx0.43.$$

现在用掷点方法来算，以 $\frac{k}{n}$ 作为 G 的近似值，这里 k 是 n 次掷点试验中，使不等式

$$\frac{1}{4}e^{\xi_i}\geqslant\eta_i\tag{7}$$

成立的 i 的个数. 由 (3)

$$D\Big(\frac{k}{n}\Big)\approx\frac{1}{n}(0.43\times0.57)=\frac{0.2451}{n}.\tag{8}$$

如果要求以 99.7% 的概率，保证以 $\frac{k}{n}$ 作为 G 的近似值时，准确到小数点后第三位，即保证

$$\Big|\frac{k}{n}-G\Big|<0.001$$

试问 n 应多大？由 (5)

$$n\geqslant\frac{9\times0.43\times0.57}{(0.001)^2}=2.2059\times10^6.\tag{9}$$

这说明试验次数相当大，亦即收敛速度较慢，这是一个缺点.

(二) **平均值方法**. 现在用另一方法来计算 (1) 中的积分，不过不必假设 $[a,b]=[0,1]$ 也不必 $0\leqslant g(x)\leqslant 1$, 只要求 $g(x)$ 在有限区间 $[a,b]$ 中可积就行了. 任取一列相互独立、同分布的随机变数 $\{\xi_i\},\xi_i$ 在 $[a,b]$ 中均匀分布，则 $\{g(\xi_i)\}$ 也是一列相互独立、同分布的随机变数，而且

$$Eg(\xi_i)=\frac{1}{b-a}\int_a^b g(x)dx=\frac{G}{b-a}. \tag{10}$$

由强大数定理

$$\lim_{n\to\infty}(b-a)\frac{\sum\limits_{i=1}^{n}g(\xi_i)}{n}=G,(\text{a.s.}).$$

因此，自然取

$$G_n=(b-a)\frac{\sum\limits_{i=1}^{n}g(\xi_i)}{n} \tag{11}$$

作为 G 的点估值，它是无偏的，即 $EG_n=G$. 如果 $g(x)$ 在 $[a,b]$ 中平方可积，则方差

$$\begin{aligned}DG_n&=\frac{(b-a)^2}{n}Dg(\xi_i)=\frac{(b-a)^2}{n}\{Eg^2(\xi_i)-[Eg(\xi_i)]^2\}\\&=\frac{1}{n}\left[(b-a)\int_a^b g^2(x)dx-\left(\int_a^b g(x)dx\right)^2\right].\end{aligned} \tag{12}$$

由此即可仿上段那样讨论精确度和计算次数的问题. 利用 §3.4 定理 1, 即得与 (5) 类似的公式

$$n\approx\frac{1}{\varepsilon^2}t_A^2\cdot(b-a)^2\cdot Dg(\xi_i). \tag{13}$$

试用此法来计算上例中的积分 (6). 由 (11)(12)

$$G_n=\frac{1}{4n}\sum_{i=1}^{n}e^{\xi_i}, \tag{14}$$

$$\begin{aligned}DG_n&=\frac{1}{n}\left[\frac{1}{16}\int_0^1 e^{2x}dx-\left(\frac{1}{4}\int_0^1 e^x dx\right)^2\right]\\&=\frac{1}{n}\left[\frac{e^2-1}{32}-\frac{(e-1)^2}{16}\right]=\frac{0.0151}{n}.\end{aligned} \tag{15}$$

如果要求的精确度与上面一样，则由 (13) 得

$$n\approx\frac{9\times 0.0151}{(0.001)^2}=0.1359\times 10^6. \tag{16}$$

将 (8) 与 (15) 、 (9) 与 (16) 分别比较，可见平均值法的估值方差小、计算次数少，因此它比掷点法好.

(12) 式还告诉我们一个降低方差的方法：如果以 $\frac{g(x)}{k}$ 代替 (12) 中的 $g(x)$, 方差就降低 k^2 倍 $(k>1)$. 从直观上看， $\frac{g(x)}{k}$ 的图形比 $g(x)$ 的振幅小 k 倍，因之方差必然降低. 所以

要计算 $\int_a^b g(x)dx$, 不如先算出 $\int_a^b \frac{g(x)}{k}dx$ 再乘以 k; 不过如计算后一积分有误差，那么这误差也会扩大 k 倍，故 k 也不能过分地大.

(三) **降低方差**. 方差小则精确度高，并可能使运算次数减少，计算速度加快，因此，降低方差是一个重要问题.

如果被积函数 $g(x)$ 可正可负，那么可把它分成为正、负两部分分别处理，故不妨设 $g(x)\geqslant 0$. 任取一密度函数 $p(x)$, 满足

$$p(x)>0,\ x\in[a,b];\quad \int_a^b p(x)dx=1, \tag{17}$$

则

$$G=\int_a^b g(x)dx=\int_a^b \frac{g(x)}{p(x)}p(x)dx=E\frac{g(x)}{p(x)},$$

其中 ξ 是以 $p(x)$ 为密度的随机变数. 这样，就把 $g(x)$ 的通常积分变为被积函数为 $\frac{g(x)}{p(x)}$ 的对密度 $p(x)$ 的积分. 设 $\{\xi_i\}$ 是一列独立的 $p(x)-$ 随机数，由上知

$$\widehat{G}_n=\frac{1}{n}\sum_{i=1}^{n}\frac{g(x)}{p(x)} \tag{18}$$

是 G 的无偏估值，有方差

$$D\widehat{G}_n=\frac{1}{n}D\,\frac{g(\xi)}{p(\xi)}=\frac{1}{n}\left[\int_a^b \frac{g^2(x)}{p(x)}dx-G^2\right]. \tag{19}$$

特别，当 $p(x)=\frac{1}{b-a}$ 为 $[a,b]$ 中均匀分布时，(18)、(19) 分别化归 (11)、(12).

现在看应如何选 $p(x)$ 使 $D\widehat{G}_n$ 最小. 如取

$$p(x)=\frac{g(x)}{\int_a^b g(x)dx} \tag{20}$$

直接代入 (19) 后得 $D\widehat{G}_n=0$. 然而实际上这是做不到的; 因为 (20) 中出现了未知的 $\int_a^b g(x)dx$, 它正是待求的. 不过如果我们事先对 $\int_a^b g(x)dx$ 有一极粗的估计, 例如 $\int_a^b g(x)dx\approx c$(常数), 那么只要取

$$p(x)\approx\frac{g(x)}{c}, \tag{21}$$

就能使方差降低.

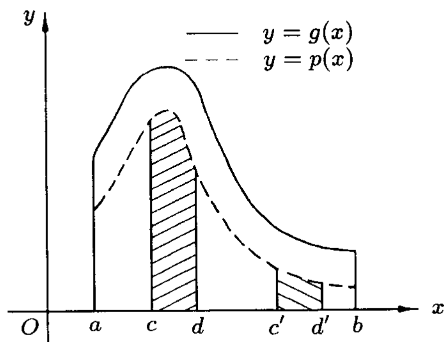


图 2

试看 (21) 的概率意义：它表示 $g(x)$ 与 $p(x)$ 的图形成比例 (图 2), $g(x)$ 大 (或小) 的地方如 $[c,d]$(或如 $[c',d']$), $p(x)$ 的值也大 (或小). 既然 $p(x)$ 是 ξ_i 的密度，ξ_i 出现在 $[c,d]$ 中的概率，就大于出现于 $[c',d']$ 中的概率. 因而 (21) 意味着：在 $g(x)$ 取大值的区间要多取一些代表点 ξ_i.

这方法还可改进：把 $[a,b]$ 分成若干子区间 $[a_j,a_{j+1}),j=1,2,\cdots,k$, 使 $g(x)$ 在每 $[a_j,a_{j+1})$ 中变化缓慢, 然后对每 $[a_j,a_{j+1})$ 施行上法. 选 $p_j(x)$ 使

$$p_j(x) \approx \frac{g(x)}{c_j}, \ (a_j \leqslant x < a_{j+1}, \ c_j \approx \int_{a_j}^{a_{j+1}} g(x)dx),$$

由于 $g(x)$ 在子区间变化缓慢， $p_j(x)$ 往往可以取为简单的函数 (如线性函数等). 取出 n_j 个 $p_j(x)$- 随机数 $\xi_{ji}(i=1,\cdots,n_j)$ 后，即得 G 的估值为

$$\overline{G}_n = \sum_{j=1}^{k} \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{g(\xi_{ji})}{p_j(\xi_{ji})}, \tag{22}$$

方差为

$$D\overline{G}_n = \sum_{j=1}^{k} \frac{1}{n_j} \left[\int_{a_j}^{a_{j+1}} \frac{g^2(x)}{p_j(x)} dx - \left(\int_{a_j}^{a_{j+1}} g(x)dx \right)^2 \right]. \tag{23}$$

当 $k=1$ 时， (22) 、 (23) 分别化为 (18) 、 (19).

以上讨论了有限积分区间 $[a,b]$ 的情形，如果是无限区间例如 $(-\infty,\infty)$, 由于假定 $\int_{-\infty}^{\infty} |g(x)|\,dx < \infty$, 故对任何 $\varepsilon > 0$, 总存在有限区间 $[a,b]$, 使

$$\left| \int_{-\infty}^{\infty} g(x)dx - \int_{a}^{b} g(x)dx \right| < \varepsilon,$$

故只要取 $[a,b]$ 充分大，即可以它代替 $(-\infty,\infty)$, 误差小于 ε.

(四) 重积分. 上述两种算法也适用于多重积分，不会发生原则性新困难，这正是 Monte-Garlo 方法的优点. 我们只简单地叙述一下平均值方法. 简记 $x = (x_1, x_2, \cdots, x_m)$ 为 m 维空间 R_m 中的点， $g(x)$ 为 m 元函数，它在 R_m 中的有限闭域 D 内可积. 今要计算

$$G \equiv \int_D g(x)dx \equiv \underset{D}{\int \cdots \int} g(x_1, \cdots, x_m)dx_1 \cdots dx_m. \tag{24}$$

取一维区间 $[a,b]$, $b-a$ 充分大，使

$$D \subset [a,b] \times [a,b] \times \cdots \times [a,b] \equiv [a,b]^m.$$

取 m 个独立、在 $[a,b]$ 中均匀分布的随机变数 $\xi_1, \xi_2, \cdots, \xi_m$, 令 $\eta = (\xi_1, \xi_2, \cdots, \xi_m)$. 我们证明：在 $\eta \in D$ 的条件下， η 在 D 中均匀分布. 换句话说，有

$$P(\eta \in V | \eta \in D) = \frac{|V|}{|D|}, \tag{25}$$

其中子域 $V \subset D, |V|$ 表 V 的 m 维体积.

实际上：

$$P(\eta \in V | \eta \in D) = \frac{P(\eta \in V)}{P(\eta \in D)}$$
$$= \frac{|V|}{(b-a)^m} \Big/ \frac{|D|}{(b-a)^m} = \frac{|V|}{|D|}.$$

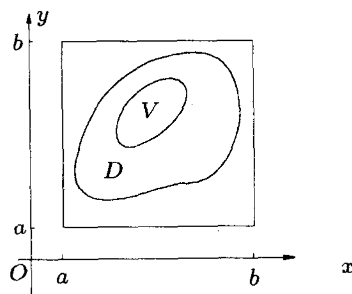


图 3

由此还推知：

$$E(g(\eta)|\eta \in D) = \frac{1}{|D|}\int_D g(x)dx = \frac{G}{|D|}.$$

根据此式，我们提出下列计算方法：象造 η 一样，造一列独立的 m 维随机向量 $\{\eta_i\}, \eta_i = (\xi_1^{(i)}, \xi_2^{(i)}, \cdots, \xi_m^{(i)})$, η_i 在 $m-$ 维立方体 $[a,b]^m$ 中有均匀分布. 如果 $\eta_1 \in D$, 就选取 η_1, 如果 $\eta_1 \bar{\in} D$, 就舍弃 η_1, 如此共选出 n 个 η_i; 记它们为 $\eta_1^0, \eta_2^0, \cdots, \eta_n^0$. 由上式得

$$|D|\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g(\eta_i^0) = G, \text{ (a.s.)},$$

于是就取

$$G_n \equiv |D|\frac{1}{n}\sum_{i=1}^{n} g(\eta_i^0) \tag{26}$$

作为 G 的无偏估值.

关于降低方差，可以采用 (三) 中类似的方法；这里还补充一点，最好采用决定性计算与随机计算相结合的方法. 在 (24) 中，可能对某些变量如 x_1, 可用数学分析方法直接算出来，这样以尽量降低积分的重数，直到剩下的变量无法直接计算时，再用平均值法. 这样所得的方差，一般会比完全用平均值法的低.

§7.2 线性方程组的解法

(一) 线性代数方程组. 设方程组为

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n, \end{pmatrix}$$

或简写为向量形式

$$X = HX + A, \tag{1}$$

其中 H, A 已知，待求的未知向量为 X. 令

$$\|H\| = \max_i\Big(\sum_j |h_{ij}|\Big).$$

我们假定 (参看 [40], 第 316 页)　$1^\circ \|H\| < 1$;

试在此条件下，用概率方法解 (1). 改写 (1) 为

$$(I - H)X = A,$$

由 1° 得

$$X = (I - H)^{-1}A = (I + H + H^2 + \cdots)A, \tag{2}$$

于是 X 的第 i 个分量 x_i 等于

$$\begin{aligned} x_i &= a_i + (HA)_i + (H^2A)_i + \cdots \\ &= a_i + \sum_{i_1} h_{ii_1} a_{i_1} + \sum_{i_1}\sum_{i_2} h_{ii_1} h_{i_1 i_2} a_{i_2} + \cdots \\ &= a_i + \sum_{k=1}^{\infty} \sum_{i_1}\sum_{i_2}\cdots\sum_{i_k} h_{ii_1} h_{i_1 i_2} \cdots h_{i_{k-1} i_k} a_{i_k}, \end{aligned} \tag{3}$$

这里 $\sum\limits_i$ 表 $\sum\limits_{i=1}^{n}$. 现在我们来定义一个马尔科夫链，使它的某一数字特征恰好等于 x_i. 任取一矩阵 $P = (p_{ij}), i, j = 1, 2, \cdots, n$, 满足条件

$$p_{ij} \geqslant 0;\ \sum_j p_{ij} < 1; \tag{4}$$

$$p_{ij} > 0,\ 如果 h_{ij} \neq 0. \tag{5}$$

令

$$p_i = 1 - \sum_j p_{ij} (\ 如 a_i \neq 0,\ 则还应取 p_i > 0). \tag{6}$$

考虑一马尔科夫链：如果它从 i 出发，那么它下一步转移到 j 的概率为 p_{ij}, 而永远停止运动的概率为 p_i, 今设它从 i 出发，经 k 步后停止运动，则运动轨道 J 可记为

$$J : i \to i_1 \to i_2 \to \cdots \to i_k,\ (k \geqslant 0, i_0 = i), \tag{7}$$

其中 $i_1, i_2, \cdots, i_k$ 为顺次经过的状态，最后停止在 i_k. 它取轨道 (7) 的概率为

$$p_{ii_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k} p_{i_k}. \tag{8}$$

在 (7) 中的轨道 J 上，定义一个函数

$$V(J) = \begin{cases} \dfrac{h_{ii_1}}{p_{ii_1}} \dfrac{h_{i_1 i_2}}{p_{i_1 i_2}} \cdot \dfrac{h_{i_{k-1} i_k}}{p_{i_{k-1} i_k}} \cdot \dfrac{a_{i_k}}{p_{i_k}}, & 如 k > 0, \\ a_i / p_i, & 如 k = 0. \end{cases} \tag{9}$$

以 E_i 表自 i 出发的条件数学期望，由 (8)(9) 得

$$\begin{aligned} E_i[V(J)] &= \frac{a_i}{p_i} \cdot p_i + \sum_{i_1} \frac{h_{ii_1} a_{i_1}}{p_{ii_1} p_{i_1}} \cdot p_{ii_1} p_{i_1} + \sum_{i_1}\sum_{i_2} \frac{h_{ii_1} h_{i_1 i_2} a_{i_2}}{p_{ii_1} p_{i_1 i_2} p_{i_2}} p_{ii_1} p_{i_1 i_2} p_{i_2} + \cdots \\ &= a_i + \sum_{k=1}^{\infty} \sum_{i_1}\sum_{i_2}\cdots\sum_{i_k} h_{ii_1} h_{i_1 i_2} \cdots h_{i_{k-1} i_k} a_{i_k}, \end{aligned}$$

由此及 (3) 即得

$$x_i = E_i[V(J)]. \tag{10}$$

于是我们得到下列计算方法：

a. 仿照 §6.3(一) 中的方法，模拟以 (p_{ij}) 为转移概率矩阵的马尔科夫链；

b. 独立地做出 n 个 (7) 中的轨道，并按 (9) 算出 $V(J_1), V(J_2), \cdots, V(J_n)$;

c. 以 $\frac{1}{n}\sum\limits_{l=1}^{n} V(J_l)$ 作为 x_i 的估值.

用这个方法来计算 x_i 所需要的转移次数设为 τ, 则

$$
\begin{aligned}
E_i\tau &= 0 \cdot p_i + 1 \cdot \sum_{i_1} p_{ii_1}p_{i_1} + 2 \cdot \sum_{i_1}\sum_{i_2} p_{ii_1}p_{i_1i_2}p_{i_2} + \cdots \\
&= \sum_{k=1}^{\infty} k \sum_{i_1}\sum_{i_2}\cdots\sum_{i_k} p_{ii_1}p_{i_1i_2}\cdots p_{i_{k-1}i_k}p_{i_k}.
\end{aligned}
\tag{11}
$$

这个方法的好处是：每个 x_i 可单独地求出，并不需要同时算出其它的 x_j.

(二) **一般解法**. 现在除去假设 1°, 来讨论一般方程组

$$
\sum_j c_{ij}x_j = d_i, \ i = 1, 2, \cdots, n \tag{12}
$$

的解. 下述解法的构思是很巧妙的. 考虑二次型

$$
Q \equiv Q(x_1, x_2, \cdots, x_n) = \sum_i \alpha_i \Big(\sum_j c_{ij}x_j - d_i\Big)^2, \tag{13}
$$

这里 $\alpha_i > 0$ 是常数，显然，求 (12) 的解等价于求 Q 的最小点 $(x_1^0, \cdots, x_n^0)$. 对于任意常数 $A > 0$,

$$
E: Q(x_1, \cdots, x_n) \leqslant A
$$

是一 n 维椭球；这只需作线性变换 $y_i = \sum\limits_j c_{ij}x_j - d_i$ 即可看出. 这个椭球的中心在 $y_i = 0 (i = 1, 2, \cdots, n)$, 也就是在 $(x_1^0, \cdots, x_n^0)$. 每个通过中心的 n 维超平面 $x_j = x_j^0$ 都把椭球分成体积相等的两部分. 这个几何性质提供了求 (12) 的解的方法：求 $x_1^0, \cdots, x_n^0$, 使椭球 E 位在 $x_j \leqslant x_j^0$ 的那部分，恰有一半的体积.

作 n 维立方体 $[a, b]^n$, 使它包含椭球 E. 设 $\xi_1, \cdots, \xi_n$ 是 n 个独立的随机变数，每个都在 $[a, b]$ 中均匀分布. 作 n 维随机向量 $\eta = (\xi_1, \cdots, \xi_n)$. 根据 §7.1(25), 得知在 "$\eta \in E$" 的条件下，$\eta$ 在椭球 E 中有均匀分布. 取 η 的 l 个独立的样本向量 $\eta_1, \cdots, \eta_l$, 如果 $\eta_1 \in E$, 就选取 η_1, 如 $\eta_1 \overline{\in} E$, 就舍弃 η_1. 如此共选出 m 个，$(m \leqslant l)$. 不妨设它们就是前面的 m 个 $\eta_1, \cdots, \eta_m$. 注意 η_i 的分量表示是

$$
\eta_i = (\xi_{i1}, \xi_{i2}, \cdots, \xi_{in}),
$$

由于 η_i 在 E 中均匀分布，而超平面 $x_j = x_j^0$ 分椭球为等体积的两部分，故当 m 充分大时，$\eta_1, \eta_2, \cdots, \eta_m$ 中，应各有一半分别在此超平面的两侧. 这表示这些向量的第 j 个分量、即 $\xi_{1j}, \xi_{2j}, \cdots, \xi_{mj}$ 之中，应约有一半不大于 (或不小于)x_j^0. 把这些分量按大小排成

$$
\xi'_{1j} \leqslant \xi'_{2j} \leqslant \cdots \leqslant \xi'_{mj}, \tag{14}
$$

其中最中间的那一个即 $\xi'_{\frac{m}{2}+1,j}$(不妨设 m 为偶数) 自然应靠近 x_j^0. 因此，我们就取

$$
\xi'_{\frac{m}{2}+1,j} \approx x_j^0, \ (j = 1, 2, \cdots, n); \tag{15}
$$

即以 $(\xi'_{\frac{m}{2}+1,1};\xi'_{\frac{m}{2}+1,2};\cdots;\xi'_{\frac{m}{2}+1,n})$ 作为方程组 (12) 的解的估值.

以上是取样本 (14) 的中数作为 x_j^0 的估值；自然也可取它的平均值作为估值，即

$$\frac{1}{m}\sum_{i=1}^{m}\xi'_{ij}\equiv\frac{1}{m}\sum_{i=1}^{m}\xi_{ij}\approx x_j^0. \tag{16}$$

(三) **积分方程的解**. 设有方程

$$x(s)=\int_a^b h(s,t)x(t)dt+a(s),\ (a\leqslant s\leqslant b), \tag{17}$$

其中 $a(s)$ 及 $h(s,t)$ 是已知的连续函数，而 $x(s)$ 是待求的未知函数. 将 (17) 与方程组 (1)，即与

$$x_i=\sum_j h_{ij}x_j+a_i,\ (i=1,2,\cdots,n)$$

比较，可见二者非常相似，只不过是从离散变量 i 过渡到连续变量 s 、把求和号 $\sum\limits_j$ 换为积分号 $\int_a^b$ 而已. 因此，可以把 (一) 中的方法移植来解 (17). 扼述如下：取二元函数 $p(s,t)$, $a\leqslant s,t\leqslant b$, 它满足 $p(s,t)\geqslant 0$, $\int_a^b p(s,t)dt\leqslant 1$, 而且若 $h(s,t)\neq 0$, 则 $p(s,t)>0$. 令 $p(s)=1-\int_a^b p(s,t)dt$. 造一马尔科夫链，如它自 s 出发，下一步转移到 $(t,t+\Delta t)$ 中的概率为 $p(s,t)\Delta t+o(\Delta t)$, 而停止运动的概率密度为 $p(s)$. 今设它自 s 出发，经 k 步后停止运动，所经轨道为 $J:s\to s_1\to s_2\to\cdots\to s_k$. 定义泛函

$$V(J)=\begin{cases}\dfrac{h(s,s_1)h(s_1,s_2)\cdots h(s_{k-1},s_k)a(s_k)}{p(s,s_1)p(s_1,s_2)\cdots p(s_{k-1},s_k)p(s_k)}, & \text{如}k>0,\\ a(s)/p(s), & \text{如}k=0.\end{cases}$$

则如 $\|h\|\equiv\sup\limits_s\int|h(s,t)|dt<1$, 就有

$$x(s)=E_s[V(J)]. \tag{18}$$

一些偏微分方程，也可用概率方法求解，请参看 §4.1 例 4. 这方面的研究很多，例如可看参考书 [27] 第 13 章. 关于蒙特卡罗方法更多的叙述可见张建中：“蒙特卡罗方法”，数学的实践与认识， 1974 年， 1,2 期.

*第八章　可靠性问题的概率分析

§8.1　可靠函数

(一) 问题的产生. 随着工农业的迅速发展，人们迫切需要提高各种设备系统的功效. 近代的设备系统都是由许多元件构成的；各种元件起着部分的作用，它们按照一定的设计方案组合起来，完成各种预定的任务. 卫星、火箭、飞机、导航系统以及各种电子仪器等等，无一不是如此. 然而，系统的构造越复杂，所含的元件越多，如果不采取适当措施，系统的工作可靠性就很可能大大降低. 这是因为，只要某些元件发生故障，就可能损害整个系统，使它不能正常工作. 为了解决这类矛盾，可靠性问题被提上了科学研究的日程. 最近若干年来，特别是第二次世界大战以后，可靠性理论得到了迅速的发展. 看来，要提高系统的可靠性，可以从两方面着手：一是提高每一组成元件的可靠性，以使系统建筑在优质元件的基础上，这样就必须对元件和系统的可靠性作依时间发展的动态分析. 二是研究系统的最佳设计、使用与维修方案，以便由可靠性较低的元件，制造出可靠性较高的大型系统，并保证它能在较长时间内正常工作. 前者是分析问题，后者是综合问题，而综合又只有在分析的基础上才能进行. 提高可靠性是一项复杂的任务，必须有工程技术人员、物理、化学等工作者共同合作，对其中许多的定量问题，还需要作深入的数学分析和计算，而概率论和数理统计正是不可缺少的数学工具. 在本章中，我们不可能对可靠性理论作全面的讨论，只是对它的分析方面作一些初步的介绍，以说明概率论在这门学科中的若干应用.

所谓可靠性，通常理解为在一段时间内无故障的概率. 对可靠性的要求，需由具体问题来决定：有时要求一段时间内无故障的概率高，而有时则要求无故障的平均时间长.

(二) 首次故障时刻. 考虑某个元件，假定它在 $t=0$ 时正常地开始工作，一直继续到时刻 $t=\tau$, 在 τ 时发生故障. 我们称 τ 为首次故障时刻，或称它为元件的寿命. 令

$$Q(t)=P(\tau\leqslant t),\ (t\geqslant 0),$$

它是 τ 的分布函数，则

$$P(t)=1-Q(t)=P(\tau>t).$$

为在 $[0,t]$ 中无故障即正常工作的概率. 称 $P(t)$ 为此元件的可靠函数. 注意 $P(0)$ 可能小于 1, 这时

$$1-P(0)=Q(0)=P(\tau=0),$$

这表示一开始就出故障的概率为 $1-P(0)>0$. 这种情形在实际中可能发生，因为元件本身可能有未被检查出的内部缺陷或损伤，或者由于保存不当、过久、运输条件不佳等而产生的新致命伤害.

一般说来，故障可分成三类：一是开始故障，它们在工作一开始或开始后不久便发生；二是随机故障，它们在元件已走上正常工作的期间，由于偶然的原因而出现；三是衰

老故障，这是由元件寿命的限制而不可避免的，它们出现在工作后期. 对一种元件，当然可能不是三种故障都全具备. 例如，饲养小猪，初出生时由于先天不足或不适应环境等原因，可能发生开始故障 (这里故障指死亡); 这一关过去，走上正常发育期，这时出现随机故障 (由于疾病或外伤等) 的机会较小；至于衰老故障在这里基本上不发生.

要具体求出某元件的寿命分布，通常有两种方法 (这对求任何分布都适用): 一是用随机抽样，通过实验，应用数理统计方法近似地求出理论分布；二是从元件的物理特性出发，提出若干基本假定，在这些假定下推导出所需分布，就象我们在 §2.4(三) 中由普阿松流推导出普阿松分布一样. 前者是归纳法，后者是演绎法.

下面我们主要利用第二种方法来导出几种常见的寿命分布. 为此，先介绍一个重要概念 - 故障率，它也可看成可靠性的一个指标.

考虑下列问题：设元件在 $[0,t]$ 中无故障，试求它在 $[t,t+\Delta t]$ 中发生故障的概率，$Q(t,t+\Delta t)$, 这里 $\Delta t>0$. 显然，我们有

$$\begin{aligned}Q(t,t+\Delta t)&=P(t+\Delta t\geqslant\tau>t|\tau>t)\\&=\frac{P(t+\Delta t\geqslant\tau>t)}{P(\tau>t)}=\frac{Q(t+\Delta t)-Q(t)}{1-Q(t)}.\end{aligned}\tag{1}$$

如果假定 $Q(t)$ 有导函数 $q(t):q(t)=Q'(t)$, 则

$$Q(t,t+\Delta t)=\frac{q(t)}{1-Q(t)}\Delta t+o(\Delta t).\tag{2}$$

我们称

$$\lambda(t)\equiv\frac{q(t)}{1-Q(t)}=-\frac{p(t)}{P(t)}\tag{3}$$

为此元件的故障率；这里 $p(t)=P'(t)$.

由 (2) 可见：$\lambda(t)$ 越大，则在 $(t,t+\Delta t]$ 中出现故障的条件概率也越大，故 $\lambda(t)$ 是表示可靠性大小的局部指标. 直观地说：如果元件在 $[0,t]$ 内正常工作，则在下一个单位时间内 (如单位长 Δt 很小) 发生故障的条件概率近似于 $\lambda(t)$.

对应于上述三个工作时期，$\lambda(t)$ 的一般图形是：在开始期 A 内，$\lambda(t)$ 下降；在正常期 B 内，由于元件工作稳定，$\lambda(t)$ 基本上是常数；在衰老期，故障增多，因之 $\lambda(t)$ 上升 (见图 1).

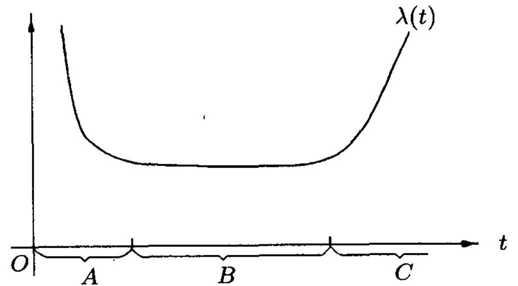


图 1

容易求出可靠函数的一般形式. 设 $p(t)$ 连续，将 (3) 两边积分，得

$$\log P(t)-\log P(0)=-\int_0^t\lambda(t)dt,$$

$$P(t)=P(0)e^{-\int_0^t\lambda(t)dt}.\tag{4}$$

由于 $\lim\limits_{t\to\infty}P(t)=0$, 故

$$\int_0^\infty\lambda(t)dt=\infty.\tag{5}$$

如果 $Q(0)=0$, 因而 $P(0)=1$, 则

$$P(t)=e^{-\int_0^t \lambda(s)ds}, \tag{6}$$

于是证明了：若可靠函数 $P(t)$ 有连续导数，则它必呈 (4) 的形式，其中 $\lambda(t)$ 是满足 (5) 的非负连续函数.

$\lambda(t)$ 可通过实验而近似求出. 设参加实验的元件共有 n 个，以 $n(t)$ 表 $[0,t]$ 内未出故障的元件数，则当 Δt 甚小而 n 很大时，由 (3) 有

$$\lambda(t)\approx\frac{P(t)-P(t+\Delta t)}{\Delta tP(t)}\approx\left(\frac{n(t)}{n}-\frac{n(t+\Delta t)}{n}\right)\Big/\Delta t\cdot\frac{n(t)}{n}=\frac{n(t)-n(t+\Delta t)}{\Delta t\cdot n(t)}.$$

换句话说：$\lambda(t)$ 近似地等于在 $[t,t+\Delta t]$ 内发生故障的元件数，除以 Δt 及在 t 以前尚未出现故障的元件数.

给出不同的 $\lambda(t)$, 就得到不同的 $P(t)$. 下面讨论几种常见的可靠函数.

(三) **常见的可靠函数.** 由于

$$Q(t)+P(t)=1,$$

所以知道 $Q(t)$ 后就知道 $P(t)$. 以下谈到分布是指寿命分布 $Q(t)$.

A) 指数分布. 这时

$$P(t)=e^{-\lambda t}, \ (t\geqslant 0);$$

$$\lambda(t)\equiv\lambda, \ (\lambda>0\text{是常数}).$$

如果某元件在任何时刻 t 的故障率都相同，即 $\lambda(t)$ 恒等于某常数 λ, 由 (6),$Q(t)$ 必有指数分布. 这使人怀疑，实际中是否有这样的元件存在，它具有不衰老的特性 (这里图 1 中只有相应于 B 的那一段). 但实际中确实是有的，至少近似地是如此. 例如：宇宙飞船的故障往往由于陨石撞击. 若设陨石在时间和空间中均匀分布，则在一段时间内飞船受击的概率不依赖于它过去是否受击，即 $\lambda(t)\equiv$ 常数，因而可认为可靠函数是指数型的. 一般地，如果某元件的故障是由一些随机碰撞所引起，而这些碰撞形成普阿松流，则它有指数型可靠函数. 此外，指数分布还可作为一些复杂系统的极限分布而出现.

B) 韦布 (Weibull,W.) 分布.

$$P(t)=e^{-\lambda t^{\alpha}}, \ (t\geqslant 0);$$

$$\lambda(t)=\alpha\lambda t^{\alpha-1}, \ (\alpha>0,\lambda>0).$$

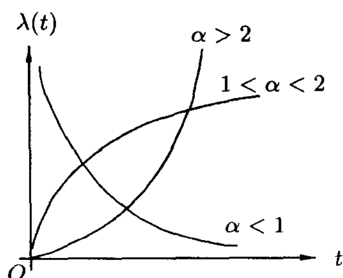


图 2

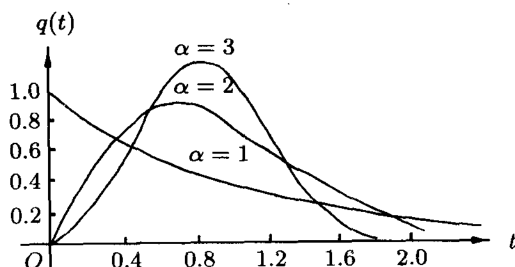


图 3

指数分布是它的特殊情形 ($\alpha=1$). 韦布分布含两个参数 $\lambda>0,\alpha>0$, 所以它比指数分布能描述更多的现象. 当 $\lambda=1$ 时, $\lambda(t)$ 及 $q(t)$ 之图见图 2 、图 3. 当 $\alpha<1$ 时, 故障率下降; $\alpha>1$ 时上升. 前者描写开始故障, 后者刻画衰老故障. 人们曾用此分布来研究电子管及滚珠轴承等的故障.

试求寿命的各级矩:

$$E(\tau^n)=\int_0^\infty \alpha\lambda t^{n-1+\alpha}e^{-\lambda t^\alpha}dt,$$

作变换 $u=t^\alpha$, 得

$$E(\tau^n)=\int_0^\infty \lambda u^{n/\alpha}e^{-\lambda u}du=\Gamma\Big(\frac{n}{\alpha}+1\Big)\lambda^{-n/\alpha}. \tag{7}$$

$\Gamma(x)$ 表 Gamma 函数. 特别,

$$E(\tau)=\Gamma\Big(\frac{1}{\alpha}+1\Big)\lambda^{-1/\alpha},$$

而方差为

$$D(\tau)=\lambda^{-2/\alpha}\Big[\Gamma\Big(\frac{2}{\alpha}+1\Big)-\Big\{\Gamma\Big(\frac{1}{\alpha}+1\Big)\Big\}^2\Big]. \tag{8}$$

C) Gamma 分布.

$$P(t)=\int_t^\infty \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x}dx,\ (t\geqslant 0);$$

$$q(t)=-P'(t)=\frac{\lambda^\alpha t^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda t},\ (\lambda>0,\alpha>0).$$

设 $X_1,\cdots,X_n$ 为独立同分布随机变数, 有参数为 λ 的指数分布, 则 $\sum\limits_{i=1}^n X_i$ 有 Gamma 分布, 参数为 λ 及 $\alpha=n$. 实际上, 此结论对 $n=1$ 显然正确; 今设它对 $n=m$ 正确, 即设 $\sum\limits_{i=1}^m X_i$ 的密度为

$$q_m(t)=\frac{\lambda^m t^{m-1}}{(m-1)!}e^{-\lambda t},$$

则 $\sum\limits_{i=1}^{m+1} X_i$ 的密度为

$$q_{m+1}(t)=\int_0^t q_m(t-x)q_1(x)dx=\frac{\lambda^{m+1}}{(m-1)!}e^{-\lambda t}\int_0^t x^{m-1}dx=\frac{\lambda^{m+1}t^m}{m!}e^-\lambda t.$$

这便证明了上述结论 (此结论也可用特征函数直接证明).

今设某元件恰好在经受 n 次撞击时出现故障, 而这些撞击相继出现于时刻 y_i, 前后两次之间的时差记为 $X_i=y_i-y_{i-1}(y_0=0)$. 设诸 X_i 相互独立, 且有相同的指数分布, 则由上面所述的结论, 可见此元件的寿命有 Gamma 分布 [1].

[1] 也可换一说法: Gamma 分布出现于如下的模型中. 设诸事故构成参数为 λ 的普阿松流. 当接连出现 $\alpha-1$ 个以上的事故时, 机器便发生故障, 则故障发生的时刻 T 有 Gamma 分布. 实际上

$$1-P(T>t)=1-\sum_{k=0}^{\alpha-1}\frac{(\lambda t)^k e^{-\lambda t}}{k!}=\int_0^t \frac{\lambda^\alpha x^{\alpha-1}}{(\alpha-1)!}e^{-\lambda x}dx,$$

参看 §2.4 的注 (见 59 页). 后一等式建立了 Gamma 分布与普阿松分布间的一个关系式.

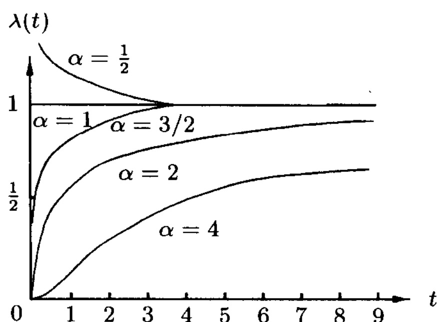


图 4

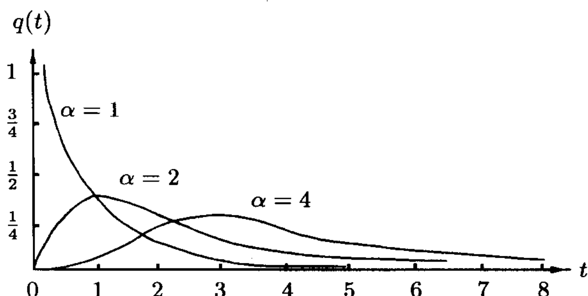


图 5

故障率 $\lambda(t) = -\frac{P'(t)}{P(t)}$, 当 $\alpha > 1$ 时，是单调上升的，有上界为 λ; 当 $\alpha < 1$ 时单调下降. $q(t)$ 的图形是单峰偏态 (非对称的). 见图 4 、图 5(其中 $\lambda = 1$).

D) 对数正态分布. 密度函数为

$$q(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\log t - a^2)}{2\sigma^2}}, \ (t > 0; \sigma > 0 \text{ 与} a \text{ 为二参数}).$$

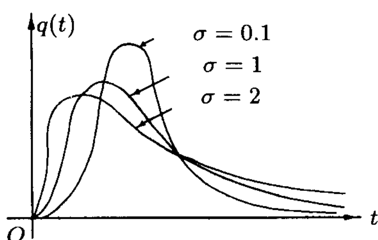


图 6

其图形见图 6. 它对应的 $\lambda(t)$ 是下降函数. 一、二级矩及方差分别为

$$E(\tau) = e^{a + \frac{\sigma^2}{2}},$$

$$E(\tau^2) = e^{2(a+\sigma^2)},$$

$$D(\tau) = e^{2a+\sigma^2}(e^{\sigma^2} - 1).$$

这分布容易由正态分布变换得来：设 X 有 $N(a, \sigma)$ 分布，令 $X = \log Y$, 则 Y 有对数正态分布.

设 $X_1, \cdots, X_n$ 为独立随机变数，取正值. 令 $Y_i = \log X_i$, 又 $Y = \sum\limits_{i=1}^{n} Y_i = \log(X_1 \cdots X_n)$. 当 $\{Y_i\}$ 满足麟德伯尔格条件时，由 §3.4 知 $\frac{Y - EY}{\sqrt{DY}}$ 有渐近 $N(0,1)$ 正态分布，因此，随机变数

$$X = X_1 X_2 \cdots X_n$$

有渐近对数正态分布. 换句话说，对数正态分布可用来描写大量非负独立变数的乘积.

E) 重指数分布. 设有厚度为 D 的金属板，它上面有许多微细小孔，第 i 个小孔的开始深度为 d_i. 由于腐蚀作用这些小孔逐渐加深，一旦有小孔穿透金属板，就发生故障. 试求首次故障时刻的分布 $Q(t)$.

我们假定 $d_i (i = 1, 2, \cdots, n)$ 相互独立，并设有相同的指数分布. 因为 $d_i \leqslant D$, 故应设有切断的指数分布，即

$$P(d_i \geqslant t) = (e^{-\lambda t} - e^{-\lambda D})/(1 - e^{-\lambda D}).$$

(这实际上是 $P(y \geqslant t | 0 \leqslant y \leqslant D)$, 这里 y 有参数为 λ 的指数分布). 设第 i 个小孔的穿透时间为 $\tau_i = k(D - d_i)$, 这里 k 为某常数，依赖于金属的性质. 则

$$F(t) \equiv P(\tau_i \leqslant t) = P\left(d_i \geqslant D - \frac{t}{k}\right) = (e^{\frac{\lambda t}{k}} - 1)/(e^{\lambda D} - 1), \ 0 \leqslant t \leqslant kD. \tag{9}$$

今以 τ 表首次故障时刻，我们有

$$\tau=\min(\tau_1,\cdots,\tau_n);$$

$$\begin{aligned}Q(t)\equiv P(\tau\leqslant t)&=1-P(\tau>t)=1-P(\tau_1>t,\cdots,\tau_n>t)\\&=1-\prod_{i=1}^{n}P(\tau_i>t)=1-[1-F(t)]^n.\end{aligned}\tag{10}$$

如果 n 非常大，则近似地有

$$Q(t)\approx 1-e^{-nF(t)},\tag{11}$$

以 (9) 代入得

$$Q(t)\approx 1-\exp\left[-\frac{n}{e^{\lambda D}-1}(e^{\frac{\lambda t}{k}}-1)\right]=1-e^{-\alpha(e^{\beta t}-1)},(t\geqslant 0),\tag{12}$$

其中常数

$$\alpha=\frac{n}{e^{\lambda D}-1},\beta=\frac{\lambda}{k}.$$

其密度为

$$q(t)=\alpha\beta e^{\beta t}e^{-\alpha(e^{\beta t}-1)}\ (t\geqslant 0)\tag{13}$$

其中， $\alpha>0,\beta>0$ 为二参数；当 $\beta=1$ 时，见图 7. 它的故障率 $\lambda(t)$ 为

$$\lambda(t)=\alpha\beta e^{\beta t}.\tag{14}$$

显然， $\lambda(t)$ 是上升函数.

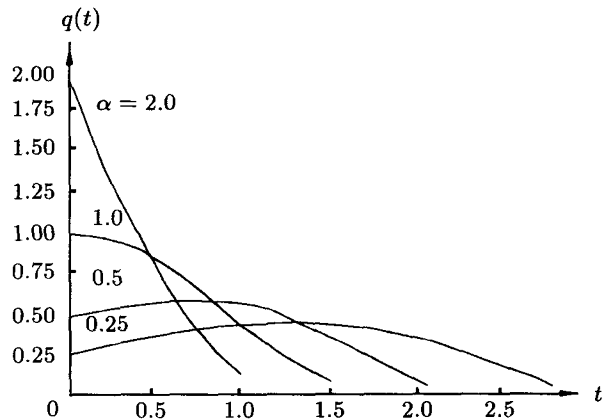


图 7

除以上所举的五种分布外，正态分布也常用作寿命分布.

我们不去列举更多的可靠函数，而只补充说明一点. 研究得最多的一类故障率是上升的 $\lambda(t)$. 通常认为这一类故障率能概括大量的现象. 另一类是下降的 $\lambda(t)$. 看来，有些问题中的 $\lambda(t)$ 可能既非上升也非下降，而是更复杂的时而上升时而下降的函数. 例如，有些系统中随时在积累能量，当能量超过一定限度时发生爆炸 (故障)，但当积累的能量未达此限度时，又可能发生能量转移或以小炸的形式释放能量. 显然，未转移以前的故障率应比转移后不久要大些，故 $\lambda(t)$ 于此时下降，此后又因积累能量， $\lambda(t)$ 又上升. 大地震的爆发可能就属于这种概型.

§8.2 更新问题

(一) 更新方程. 上节讨论了首次故障以前的概率性质，现在来研究在此以后的问题．元件出了故障，就需要更新 (修理，或者换上新的同样的元件)，我们先假定更新所需的时间为 0. 第二个元件到时候也需更新，如此继续下去．于是得到一列时刻 $\{t_n\}$:

$$0 \equiv t_0 < t_1 < t_2 < \cdots < t_n < \cdots,$$

t_n 表第 n 次更新时刻，而 $\tau_n = t_n - t_{n-1}$ 是第 n 个元件的寿命．我们假定 $\{\tau_n\}$ 是一列独立随机变数，有相同的连续分布函数 $Q(t) = P(\tau_n \leqslant t)$.

称随机数列 $\{t_n\}$ 为更新过程；当 $Q(t)$ 为指数分布时，它化为普阿松流．

对每个固定的 $t > 0$, 定义随机变数 $v(t)$:

$$t_{v(t)} < t \leqslant t_{v(t)+1}; \tag{1}$$

换句话说： $v(t)$ 是在 t 以前的更新次数．显然 $v(t)$, 只能取非负整数为值．

试求 $v(t)$ 的分布．为此，先求 t_n 的分布 $F_n(t)$. 由于

$$t_n = \tau_1 + \tau_2 + \cdots + \tau_n$$

是 n 个相互独立同分布的随机变数的和，由 §2.8(5)

$$F_n(t) = \int_0^t F_{n-1}(t-s)dQ(s),\ F_1(t) = Q(t), \tag{2}$$

于是得

$$\begin{aligned} P(v(t) = n) &= P(v(t) \geqslant n) - P(v(t) \geqslant n+1) \\ &= P(t_n < t) - P(t_{n+1} < t) = F_n(t) - F_{n+1}(t); \end{aligned} \tag{3}$$

$$P(v(t) = 0) = 1 - Q(t).$$

在更新问题的研究中，起重要作用的是更新函数 $M(t)$, 它是 $v(t)$ 的数学期望：

$$M(t) = Ev(t). \tag{4}$$

由 (3), 容易求得 [1]

$$M(t) = \sum_{n=0}^{\infty} nP(v(t) = n) = \sum_{n=1}^{\infty} F_n(t). \tag{5}$$

---

[1] 如 $Q(t) < 1$, 则 $M(t) < \infty$. 此因 $F_{n-1}(t)$ 单调上升，由 (2), 知 $F_n(t) \leqslant [Q(t)]^n$, 故 $M(t) = \sum\limits_{n=1}^{\infty} F_n(t) \leqslant \sum\limits_{n=1}^{\infty} [Q(t)]^n < \infty$.

(5) 中右方的级数实际上不容易计算，但我们可以通过解方程来求 $M(t)$: 由 (5)(4),

$$M(t) = Q(t) + \sum_{n=2}^{\infty} \int_0^t F_{n-1}(t-s)dQ(s)$$
$$= Q(t) + \int_0^t \sum_{n=1}^{\infty} F_n(t-s)dQ(s) = Q(t) + \int_0^t M(t-s)dQ(s),$$

因此， $M(t)$ 是方程

$$M(t) = Q(t) + \int_0^t M(t-s)dQ(s) \tag{6}$$

的解. 一般地，人们称下列积分方程为更新方程：

$$M(t) = G(t) + \int_0^t M(t-s)dQ(s), \tag{7}$$

其中 $G(t), Q(t), (t \geqslant 0)$ 是已知函数，而 $M(t)$ 是未知的. (6) 是 (7) 在 $G(t) = Q(t)$ 时的特殊情形.

如果存在导数

$$q(t) = Q'(t),\ m(t) = M'(t),\ g(t) = G'(t),$$

微分 (7) 得

$$m(t) = g(t) + \int_0^t m(t-s)q(s)ds. \tag{8}$$

称 $m(t)$ 为更新密度，由 (5), 在可逐项微分的条件下，有

$$m(t) = \sum_{n=1}^{\infty} f_n(t),\ f_n(t) = F_n'(t). \tag{9}$$

对 (6) 式两边取拉普拉斯 (Laplace) 变换，由于卷积的拉普拉斯变换等于二函数的拉普拉斯变换的积，故

$$\overline{M}(s) = \overline{Q}(s) + \overline{M}(s)\overline{Q}(s),$$

或

$$\overline{M}(s) = \frac{\overline{Q}(s)}{1-\overline{Q}(s)}, \tag{10}$$

这里 $\overline{M}(s) = \int_0^{\infty} e^{-st}dM(t)$, 又

$$\overline{Q}(s) = \int_0^{\infty} e^{-st}dQ(t) = E(e^{-s\tau}),$$

τ 为首次故障时刻. 由 (10), 知 $\overline{M}(s)$ 与 $\overline{Q}(s)$ 相互决定，故 $M(t)$ 与 $Q(t)$ 也相互决定. 既然 $Q(t)$ 决定了更新过程的一切概率性质，所以 $M(t)$ 也能起此作用. 这就是为什么 $M(t)$ 特别重要的原因.

例 1 设 $Q(t)$ 为 Gamma 分布，密度为

$$q(t) = \frac{\lambda^m t^{m-1}}{(m-1)!}e^{-\lambda t},\ (t \geqslant 0),$$

根据 §8.1(三) 中关于 Gamma 分布的一个结论，可见 $F_n(t)$ 可看成为 nm 个指数分布的卷积，故它也是 Gamma 分布，密度为

$$f_n(t) = \frac{\lambda^{nm} t^{nm-1}}{(nm-1)!} e^{-\lambda t}. \tag{11}$$

由此得

$$F_n(t) = \int_0^t f_n(s) ds = 1 - e^{-\lambda t} \sum_{r=0}^{nm-1} \frac{(\lambda t)^r}{r!}, \ (n \geqslant 1). \tag{12}$$

从而

$$\begin{aligned} P(v(t) = n) &= F_n(t) - F_{n+1}(t) \\ &= e^{-\lambda t} \left(\sum_{r=0}^{nm+m-1} - \sum_{r=0}^{nm-1} \right) \frac{(\lambda t)^r}{r!} = e^{-\lambda t} \sum_{r=nm}^{nm+m-1} \frac{(\lambda t)^r}{r!}. \end{aligned}$$

由 (9)(1)

$$m(t) = \sum_{n=1}^{\infty} \frac{\lambda^{nm} t^{nm-1}}{(nm-1)!} e^{-\lambda t}.$$

特别 [1], 如 $m = 1$, 即当 $Q(t)$ 为指数分布时，

$$m(t) = \lambda, \ M(t) = \lambda t.$$

如 $m = 2$, 则

$$m(t) = \lambda/2 - (\lambda/2) e^{-2\lambda t}, \ M(t) = \lambda t/2 - \frac{1}{4} + \frac{1}{4} e^{-2\lambda t}.$$

回到一般的情形，试讨论下列问题：应该储存多少元件，使以 99% 以上的概率，在 t_0 以前足够用？这就是说，要求最小的正整数 n, 使

$$P(v(t) < n) \geqslant 99\%; \ 或 P(v(t) \geqslant n) < 1\%.$$

由 (12)

$$P(v(t) \geqslant n) = F_n(t) = 1 - e^{-\lambda t} \sum_{r=0}^{nm-1} \frac{(\lambda t)^r}{r!} < 1\%.$$

解此不等式，即可求出最小的 n; 这里的求和有表可查.

(二) 若干极限定理.

定理 1 设 $E(\tau_n) = \mu, D(\tau_n) = \sigma^2 < \infty$. 则当 $t \to \infty$ 时， $v(t)$ 有渐近 $\left(\frac{t}{\mu}, \sqrt{\frac{t\sigma^2}{\mu^3}}\right)$ 正态分布；即

$$\lim_{t \to \infty} P\left(\frac{v(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy. \tag{13}$$

---

[1] 对一般的 m, 有 $M(t) = \frac{\lambda t}{m} + \frac{1}{m} \sum_{j=1}^{m-1} \frac{\theta^j}{1-\theta^j} [1 - e^{-\lambda t (1-\theta^j)}]$, 其中 $\theta = e^{(2\pi t/m)}$. 参看 [39],p.175-177.

证 令 t,n 保持下列关系

$$\frac{t-n\mu}{\sigma\sqrt{n}}=-x,\quad (x\text{ 固定}), \tag{14}$$

而趋于 ∞, 即 $t\to\infty, n\to\infty$. 记 $t_n=\sum\limits_{i=1}^{n}\tau_i$, 有

$$\begin{aligned}P(v(t)<n)&=P(t_n>t)=P(t_n>n\mu-x\sigma\sqrt{n})\\&=P\Big(\frac{t_n-n\mu}{\sigma\sqrt{n}}>-x\Big)\to\frac{1}{\sqrt{2\pi}}\int_{-x}^{\infty}e^{-\frac{y^2}{2}}dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy.\end{aligned} \tag{15}$$

另方面,

$$\begin{aligned}P(v(t)<n)&=P\Big(\frac{v(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}}<\frac{n-t/\mu}{\sqrt{t\sigma^2/\mu^3}}\Big)\\&=P\Big(\frac{v(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}}<x\Big[1+\frac{x\sigma\sqrt{n}}{t}\Big]^{\frac{1}{2}}\Big),\end{aligned} \tag{16}$$

这里用到 (14). 再由 (14)

$$\mu n-x\sigma\sqrt{n}-t=0.$$

把此式看成 $\sqrt{n}$ 的二次方程，解得

$$\sqrt{n}=\frac{x\sigma+\sqrt{(x\sigma)^2+4t\mu}}{2\mu}.$$

故 $\lim\limits_{t\to\infty}\frac{\sqrt{n}}{t}=0$. 由此及 (15)(16) 即得 (13). □

例 2 设某类元件的平均寿命为 $\mu=150$ 小时，方差 $\sigma^2=2500$ 小时. 今欲工作 3600 小时，问需储备多少元件，才能以 95% 的概率保证够用. 这问题与例 1 同，不过这里的近似解法中无需知道寿命的分布律，只需知道它的平均值与方差. 先解

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy=95\%,$$

查表得 $x\approx1.65$. 由 (13) 知

$$p\Big(v(t)<1.65\sqrt{\frac{t\sigma^2}{\mu^3}}+\frac{t}{\mu}\Big)\approx95\%.$$

以题中所给的数据代入，得

$$n\approx1.65\sqrt{\frac{3600\cdot2500}{150^3}}+\frac{3600}{150}\approx27.\ \square$$

下列结果的直观意义非常明显，而证明却相当烦难，故只叙述而不证明.

1)

$$\lim_{t\to\infty}\frac{M(t)}{t}=\frac{1}{\mu},$$

(如 $\mu=\infty$, 则右方理解为 0).

这表明：在 $[0,t]$ 中的平均更新次数乘平均寿命之积渐近地等于 t.

2) 设 $Q(t)$ 连续，则

$$\lim_{t\to\infty}[M(t+h)-M(t)]=\frac{h}{\mu},\tag{17}$$

其中 $h>0$ 任意. 如果 $Q(t)$ 是格子点型的离散分布，即如它的密度为 $\begin{pmatrix} nd & \\ qn; & n=1,2,\cdots \end{pmatrix}$, $d>0$ 为步长，则 (17) 当 $h=nd$ $(n=1,2,\cdots)$ 时仍正确. 无论在何种情形，对 $\mu=\infty$ 结论仍正确，但应理解 (17) 右方为 0.

(17) 的直观意义为：在 $[t,t+h)$ 中的平均更新次数与平均寿命之积渐近地等于 h.

3) 设 $Q(t)$ 连续，又 $g(t)$ 是 $[0,\infty)$ 中有界变差函数，则

$$\lim_{t\to\infty}\int_0^t g(t-s)dM(s)=\frac{1}{\mu}\int_0^\infty g(s)ds,\tag{18}$$

只要右方积分存在 (如 $\mu=\infty$, 则理解右方为 0).

4) 如果 τ_i 的方差 σ^2 有限，则

$$\lim_{t\to\infty}\left[M(t)-\frac{t}{\mu}\right]=\frac{\sigma^2}{2\mu^2}-\frac{1}{2}.$$

换言之，当 t 充分大时

$$M(t)\approx\frac{t}{\mu}+\frac{\sigma^2}{2\mu^2}-\frac{1}{2}.$$

5) 如 $Q(t)$ 有连续密度 $q(t)$, 而且 $\lim\limits_{t\to\infty}q(t)=0$, 则

$$\lim_{t\to\infty}m(t)=\frac{1}{\mu},\ (m(t)=M'(t)).$$

(三) **剩余寿命的分布**. 对任意固定的 $t>0$, 以 ξ_t 表自 t 算起直到下次故障的时长，即

$$\xi_t=t_{v(t)+1}-t,$$

称 ξ_t 为在时刻 t 的剩余寿命. 试求

$$P_t(\tau)\equiv P(\xi_t>\tau),$$

它是在 $[t,t+\tau]$ 中正常工作的概率. 为此，考虑事件

$$A_0=\{t+\tau<s_1\},A_n=\{t_n<t<t+\tau<t_n+\tau_{n+1}\},(n\geqslant 1).$$

这些事件互不相容，而且

$$(\xi_t>\tau)=\bigcup_{n=0}^{\infty}A_n,$$

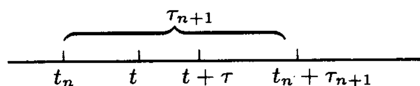


图 2

故 $P_t(\tau)=\sum\limits_{n=0}^{\infty}P(A_n)$.

但 $P(A_0)=1-Q(t+\tau)$; 又因 $\{\tau_n\}$ 有相同分布 $Q(t)$, 故

$$P(A_n)=\int_0^t P(x\leqslant t_n\leqslant x+dx)P(\tau_{n+1}>t+\tau-x)$$
$$=\int_0^t[1-Q(t+\tau-x)]f_n(x)dx,$$

因此，由 (9) 得

$$P_t(\tau)=1-Q(t+\tau)+\int_0^t[1-Q(t+\tau-x)]m(x)dx. \tag{19}$$

现在考虑当 t 充分大以后的稳定情形. 在 (19) 中令 $t\to\infty$, 得

$$P(\tau)\equiv\lim_{t\to\infty}P_t(\tau)=\lim_{t\to\infty}\int_)^t[1-Q(t+\tau-x)]m(x)dx,$$

利用 (18), 取那里的 $g(t)=1-Q(t+\tau)$, 即得

$$P(\tau)=\frac{1}{\mu}\int_0^\infty[1-Q(t+\tau)]dt=\frac{1}{\mu}\int_\tau^\infty[1-Q(t)]dt. \tag{20}$$

回忆 $\mu=E(\tau_n)$, 由 §2.9(七), $\mu=\int_0^\infty[1-Q(x)]dx$, 故

$$P(\tau)=\frac{\int_\tau^\infty[1-Q(t)dt}{\int_0^\infty[1-Q(t)dt},$$

这式的几何意义非常明显.

(四) 更新时间大于 0 的情形. 在以上的讨论中，都假定故障是立即排除的，现在考虑一般情形，即更换一个元件需要停工一段时间，等更新完毕后才能继续工作. 这就是说：工作到时间 τ_1', 发生故障，由于更新，需要一段时间 τ_1''; 然后再工作一段时间 τ_2', 又需更新一段时间 τ_2''; $\cdots$. 我们假定：

(i) $\tau_1',\tau_1'',\tau_2',\tau_2''\cdots$ 是相互独立的随机变数；

(ii) $\{\tau_n'\}$ 有相同的分布 $Q_1(t)$; $\{\tau_n''\}$; 也有相同的分布函数 $Q_2(t)$.

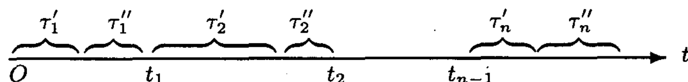


图 3

令 $\tau_n=\tau_n'+\tau_n''$, 则 $\{\tau_n\}$ 独立，有相同分布为

$$Q(t)\equiv P(\tau_n\leqslant t)=\int_0^t Q_1(t-s)dQ_2(s).$$

于是可把 $\{t_n\}$ 视为更新时长为 0 的更新过程，这里 $t_n=\sum_{i=1}^n\tau_i$, 因之本节 (一) 、(二) 、(三) 中的一些结果可用于 $\{t_n\}$, 不过要注意这时 $Q(t)$ 是 $Q_1(t)$ 与 $Q_2(t)$ 的卷积. 例如，以 $F_n(t)$ 表 $Q(t)$ 的 n 重卷积，则更新函数为

$$M(t)=\sum_{n=1}^\infty F_n(t),$$

更新密度为

$$m(t)=\sum_{n=1}^{\infty}f_n(t),\ f_n(t)=F_n'(t).$$

由 (二)(5),

$$\lim_{t\to\infty}m(t)=\frac{1}{\mu_1+\mu_2},$$

这里 $\mu_1=E(\tau_n'),\mu_2=E(\tau_n'')$. 如以 $v(t)$ 表在 t 以前的更新次数，则由定理 1,$v(t)$ 有渐近 $N(a,\sigma)$ 正态分布，其中

$$a=\frac{t}{\mu_1+\mu_2},\quad \sigma^2=\frac{(\sigma_1^2+\sigma_2^2)t}{(\mu_1+\mu_2)^3},$$

σ_1^2,σ_2^2 分别表 τ_n',τ_n'' 的方差.

以 $P_t(\tau)$ 表在区间 $[t,t+\tau]$ 中正常工作 (无故障) 的概率. 完全象推导 (19)(20) 一样，得

$$P_t(\tau)=1-Q_1(t+\tau)+\int_0^t[1-Q_1(t+\tau-x)]m(x)dx; \tag{21}$$

$$P(\tau)\equiv\lim_{t\to\infty}P_t(\tau)=\frac{1}{\mu_1+\mu_2}\int_\tau^\infty[1-Q_1(t)]dt. \tag{22}$$

令 $A=\lim\limits_{\tau\to0}P(\tau)$, 则由 (22)

$$A=\frac{1}{\mu_1+\mu_2}\int_0^\infty[1-Q_1(t)]dt=\frac{\mu_1}{\mu_1+\mu_2}. \tag{23}$$

称 A 为工作系数，它表示当 t 充分大时，在时刻 t 元件正常工作的概率近似于 A. (23) 的直观意义是明显的：右方是元件的平均寿命 (亦即平均工作时间) 与总平均时间 (平均工作时间加平均更新时间) 的比. 利用 A, 还可把 (22) 改写成

$$P(\tau)=A\cdot\frac{1}{\mu_1}\int_\tau^\infty[1-Q_1(t)]dt. \tag{24}$$

将此式与 (20) 对照，可见在长为 τ 的区间中正常工作的概率，等于在此区间开始时正常工作的概率，乘以在更新时间为 0 的系统中，在长为 τ 的区间中正常工作的概率.

§8.3　系统的可靠性

(一) **串联系统**. 系统是由许多元件按一定的方式组合起来的. 因此，系统的可靠性，既依赖于每个元件的可靠性，也依赖于元件之间的联接关系. 我们的目的是通过元件的可靠性来求出系统的可靠性. 一种最简单的关系是串联，这时只要有一个元件发生故障，就使整个系统发生故障 (见图 1). 多级火箭是一个很好的例子：第一级正常工作后，必须点燃第二级，使它也正常工作，然后自行脱落；以后各级顺次如此. 如果有一级出故障，就使整个火箭的运用发生问题. 这个例子中的工作程序是按时间先后进行的，对一般的串联系统可不必如此.

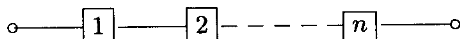


图 1

以 A_t(或 $A_{m,t}$) 表事件: "在时刻 t, 系统 (或第 m 个元件) 正常工作", 在串联情况下, 当且只当一切 $A_{m,t}, (m=1,2,\cdots,n; n$ 为元件个数) 出现时 A 出现, 故

$$A_t = A_{1,t}A_{2,t}\cdots A_{n,t} \tag{1}$$

$$\begin{aligned} P(A_t) &= P(A_{1,t}A_{2,t}\cdots A_{n,t}) \\ &= P(A_{1,t})P(A_{2,t}|A_{1,t})P(A_{3,t}|A_{1,t}A_{2,t})\cdots P(A_{n,t}|A_{1,t}\cdots A_{n-1,t}). \end{aligned} \tag{2}$$

现在设备元件的可靠性是相互独立的, 这就是说, 任何一些元件发生故障并不影响其它元件是否发生故障; 用数学语言来表达, 就是 $A_{1,t}\cdots A_{n,t}$ 是独立事件. 这时 (2) 式简化为

$$P(A_t) = P(A_{1,t})P(A_{2,t})\cdots P(A_{n,t}).$$

以 $P(t)$(或 $p_m(t)$) 表系统 (或第 m 个元件) 的可靠函数, 可改写上式为

$$P(t) = p_1(t)p_2(t)\cdots p_n(t). \tag{3}$$

设 $p_m(0)=1$, 第 m 个元件 (或系统) 的事故率记为 $\lambda_m(t)$(或 $\lambda(t)$), 则由 §8.1(6) 及上式得

$$e^{-\int_0^t \lambda(s)ds} = e^{-\int_0^t \lambda_1(s)ds-\cdots-\int_0^t \lambda_n(s)ds},$$

从而

$$\lambda(t) = \lambda_1(t)+\lambda_2(t)+\cdots+\lambda_n(t). \tag{4}$$

特别, 当每个元件的寿命有指数分布时, $\lambda_m(t)\equiv\lambda_m$, 故 $\lambda = \sum_{m=1}^{n}\lambda_m$. 既然系统的寿命有指数分布 $1-e^{-\lambda t}$, 系统的平均寿命 μ 就是

$$\mu = \frac{1}{\lambda} = \frac{1}{\lambda_1+\cdots+\lambda_n} = \frac{1}{\frac{1}{\mu_1}+\cdots+\frac{1}{\mu_n}}, \tag{5}$$

这里 μ_m 为第 m 个元件的平均寿命.

(二) **并联系统**. 另一种元件关系是并联, 这时当且仅当一切元件发生故障时系统发生故障. (见图 2). 当元件的可靠性相互独立时, 由并联的定义得

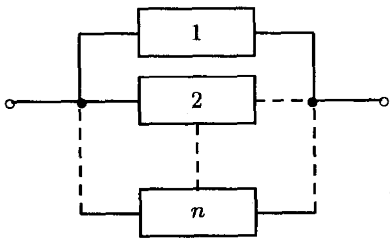


图 2

$$1-P(t) = \prod_{m=1}^{n}[1-p_m(t)], \tag{6}$$

亦即

$$P(t) = 1-\prod_{m=1}^{n}[1-p_m(t)]. \tag{7}$$

特别, 若 $p_m(t)=e^{-\lambda t}, (m=1,2,\cdots,n)$, 得

$$P(t) = 1-(1-e^{-\lambda t})^n. \tag{8}$$

这时系统的平均寿命为 μ:

$$\mu = \int_0^\infty P(t)dt = \int_0^\infty [1-(1-e^{-\lambda t})^n]dt.$$

作变换 $1-e^{-\lambda t}=x$, 得

$$\begin{aligned}\mu &= \frac{1}{\lambda}\int_0^1 \frac{1-x^n}{1-x}dx = \frac{1}{\lambda}\int_0^1 (1+x+\cdots+x^{n-1})dx \\ &= \frac{1}{\lambda}\Big(1+\frac{1}{2}+\cdots+\frac{1}{n}\Big),\end{aligned}$$

对充分大的 n 有

$$\mu \approx \frac{1}{\lambda}(\log n + c). \quad c = 0.57712\cdot(c\ 为欧拉常数) \tag{9}$$

在一般的系统中，常是既有串联、又有并联. 例如，设某系统由 N 个子系统串联而成，而每个子系统又由 k_j 个相同的元件并联组成 (见图 3). 设第 j 类元件的可靠函数为 $p_j(t)$, 令 $q_j(t)=1-p_j(t)$. 由 (7), 第 j 个子系统的可靠函数为 $1-q_j^{k_j}(t)$, 再由 (3) 知整个系统的可靠函数为

$$P(t) = \prod_{j=1}^N [1-q_j^{k_j}(t)].$$

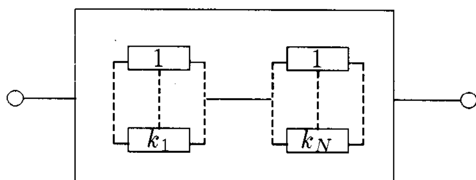


图 3

例 1 设有 n 个相同的独立元件并联如图 2, 问 n 至少有多大，才能使系统在时刻 t 的可靠性达到 95%?

解 设元件的可靠函数为 $p(t)$, 由 (7) 得

$$[1-p(t)]^n \leqslant \frac{5}{100},$$

故

$$n \geqslant \Big(\log\frac{5}{100}\Big)\Big/\log[1-p(t)].$$

例 2 储备元件的最佳分配. 设系统的组成如图 3 所示. 每个子系统中，只有一个元件起基本作用，其余起储备作用，但它们都处于相同状态. 一旦基本元件出故障，储备元件中的某个立即补上. 显然，储备元件越多，系统可靠性越高，但所需费用也越大，如何处理这对矛盾？

设第 j 类元件的单价为 c_j. 数学上这个问题有两种提法：

1) 在总费用不超过某数 c 的条件下，决定每类元件的个数，使系统的可靠性极大. 这也就是说：试求正数 $k_1,k_2,\cdots,k_N$, 使满足条件

$$\sum_{j=1}^N k_jc_j \leqslant c,$$

并使

$$P(t)=\prod_{j=1}^{N}[1-q_j^{k_j}(t)] \tag{10}$$

达到极大值. 这个问题可用动态规划求解，我们不去叙述它的解法 [1] 只给出一组近似解：

$$k_j\approx\left[\frac{c}{\log q_j(t)}\right]\Big/\left[\sum_{j=1}^{N}\frac{c_j}{\log q_j(t)}\right],\quad (j=1,\cdots,N)$$

2) 在希望系统的可靠系数不小于某定数 P 的条件下，求出各类元件的个数，使总费用极小. 这也就是说：试求正数 $k_1,k_2,\cdots,k_N$, 使满足条件

$$\prod_{j=1}^{N}[1-q_j^{k_j}(t)]\geqslant P. \tag{11}$$

并使

$$c=\sum_{j=1}^{N}k_jc_j$$

达到极小值. 可以证明：近似解为

$$k_j\approx\frac{1}{\log q_j(t)}\log\frac{c_j(1-P)/\log q_j(t)}{\sum\limits_{j=1}^{N}c_j/\log q_j(t)}.$$

注. 设系统（或第 j 个元件）的寿命为 τ(或为 τ_j), 则对串联系统, $\tau=\min(\tau_1,\cdots,\tau_n)$; 而对并联系统, $\tau=\max(\tau_1,\cdots,\tau_n)$. 因而串联与并联间有对偶关系. 例如, 在 (3) 中以 $Q(t)$ 替换 $P(t)$, 就得到对并联成立的公式 (6).

(三) **纯灭模型**. 以上讨论中假定了元件间的可靠性是独立的，即某些元件出故障并不影响其它元件出故障的概率. 但在实际问题中并不常如此. 一般地，某些元件发生故障后，由于工作元件减少，因而平均负担加大，剩下的元件出故障的概率就增加. 这意味着每个元件发生故障的概率与工作元件的个数有关. 下面来考虑一个系统，它由 n 个元件并联而成；我们假定：在任一时刻，如有 k 个元件工作，则每一元件在此时刻的故障率等于同一常数 λ_k, 它不依赖于时间 t 而只依赖于工作元件的个数 k.

以 x_t 表在时刻 t 正常工作的元件个数，它是随机变数，可能取的值是 $0,1,2,\cdots,n$. 这里的随机过程 $\{x_t,t\geqslant 0\}$ 属于所谓纯灭过程一类，它的转移情况是 $k\to k-1\to k-2\to\cdots\to 0$. 由于系统是并联的，当且只当 $x_t=0$ 时，系统在 t 时发生故障. 故

$$P(t)=1-P(x_t=0). \tag{12}$$

我们说系统在 t 时处于状态 k, 如果 $x_t=k$. 令

$$p_k(t)=P(x_t=k).$$

[1]参看 Полляк Д. Г., “О двух задачах теорин надежности радиоэлектронного оборудования” Радиотехника, 15, No.10, 1960.

为了求出可靠函数 $P(t)$, 由 (12), 只要求出 $p_0(t)$. 试推导出 $p_k(t)$ 所满足的微分方程.

设在 t 时系统处于状态 k, 由假定，在很短时间 Δt 内，系统以概率 $k\lambda_k\Delta t + o(\Delta t)$ 转移到状态 $k-1$, 或以概率 $1-k\lambda_k\Delta t + o(\Delta t)$ 停留在 k, 而转移到其它状态的概率则是高级无穷小 $o(t)$. 由全概率公式，得

$$\begin{aligned}p_k(t+\Delta t) = &[(k+1)\lambda_{k+1}\Delta t + o(\Delta t)]p_{k+1}(t)\\ &+ [1-(k\lambda_k\Delta t + o(\Delta t))]p_k(t) + o(\Delta t).\end{aligned} \tag{13}$$

以 Δt 除两边，令 $\Delta t \to 0$, 得

$$p_k'(t) = (k+1)\lambda_{k+1}p_{k+1}(t) - k\lambda_k p_k(t),\ (k=0,1,\cdots,n-1). \tag{14}$$

当 $k=n$ 时， (13) 中右边缺少第一项，故

$$p_n'(t) = -n\lambda_n p_n(t). \tag{15}$$

由 (15) 及开始条件

$$p_n(0)=1,\ p_k(0)=0,\ (k<n) \tag{16}$$

可解出 $p_n(t)$, 代入 (14) 最后一式可求出 $p_{n-1}(t)$, 如此继续可求出一切 $p_k(t)$. 但不如用拉普拉斯变换求解来得简便，令

$$a_k(z) = \int_0^\infty p_k(t)e^{-zt}dt,\ (k=0,1,\cdots,n).$$

以 e^{-zt} 乘 (14)(15) 两边，对 t 自 0 到 ∞ 积分，由拉普拉斯变换的性质，得代数方程组

$$za_n(z) - 1 = -n\lambda_n a_n(z),$$

$$za_k(z) = (k+1)\lambda_{k+1}a_{k+1}(z) - k\lambda_k a_k(z),\ (k<n).$$

解得

$$a_n(z) = \frac{1}{z+n\lambda_n},$$

$$a_k(z) = \frac{(k+1)\lambda_{k+1}\cdots n\lambda_n}{(z+k\lambda_k)\cdots(z+n\lambda_n)}.$$

特别，

$$a_0(z) = \frac{n!\lambda_1\cdots\lambda_n}{z(z+\lambda_1)(z+2\lambda_2)\cdots(z+n\lambda_n)}.$$

取 $a_0(z)$ 的反拉普接斯变换，查表即可得

$$p_0(t) = 1 - n!\lambda_1\cdots\lambda_n\sum_{k=1}^{n}\frac{e^{-k\lambda_k t}}{k\lambda_k b'(-k\lambda_k)},$$

其中

$$b(x) = (x+\lambda_1)(x+2\lambda_2)\cdots(x+n\lambda_n),$$

$$b'(-k\lambda_k) = \left.\frac{db(x)}{dx}\right|_{x=-k\lambda_k}.$$

这样就可求得系统的可靠函数 $P(t) = 1 - p_0(t)$.

系统的平均寿命 μ(即系统的首次故障时刻的数学期望) 容易求出：定义

$$\mu(z) \equiv \int_0^\infty P(t)e^{-zt}dt = \int_0^\infty [1 - p_0(t)]e^{-zt}dt = \frac{1}{z} - a_0(z),$$

令 $z \to 0$, 得

$$\mu(0) = \mu = \int_0^\infty P(t)dt = \frac{1}{n\lambda_n} + \frac{1}{(n-1)\lambda_{n-1}} + \cdots + \frac{1}{\lambda_1}.$$

(四) 更新系统. 现在考虑首次故障以后的情形. 设系统由 n 个元件构成，我们假定，任何一些元件的故障与更新并不影响其它元件的可靠性. 对于第一个元件，有一列故障时刻；对第二个元件，$\cdots$, 第 n 个元件也都如此，这 n 列故障的总体构成系统的故障列 (见图 4). 这里我们假定元件的更新是瞬时完成的，即更新时间为 0. 以 $v(t)$(或 $v_k(t)$) 表系统 (或第 k 个元件) 在时刻 t 以前的更新次数，$M(t) = Ev(t), M_k(t) = Ev_k(t)$. 则

$$v(t) = \sum_{k=1}^n v_k(t),$$

$$M(t) = \sum_{k=1}^n M_k(t).$$

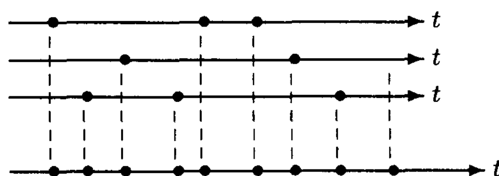


图 4

系统的更新密度 $m(t) = M'(t)$ 与元件的更新密度 $m_k(t)$ 间的关系为

$$m(t) = \sum_{k=1}^n m_k(t).$$

其次，由所假定的元件间的独立性，显然有

$$P(v(t) = r) = \sum_{r_1+r_2+\cdots+r_n=r} P(v_1(t) = r_1)P(v_2(t) = r_2)\cdots P(v_n(t) = r_n).$$

再次，试求在 $[t, t+\tau]$ 中系统正常工作 (即无故障) 的概率 $P_t(\tau)$. 这事件当且仅当每一元件都在 $[t, t+\tau]$ 中无故障时发生. 由独立性假定得

$$P_t(\tau) = P_{1,t}(\tau)P_{2,t}(\tau)\cdots P_{n,t}(\tau),$$

这里 $P_{k,t}(\tau)$ 是对第 k 个元件同样事件的概率，它可由 (§8.2 (19) 求出.

现在来考虑更新时间大于 0 的情形. 这时重要的可靠性指标为系统的工作系数 A, 它是当系统处于平稳时、也就是当 t 充分大时，在时刻 t 系统正常工作的概率的极限值. 我们仍假定上述元件间的独立性. 由于当且仅当每一元件都正常工作时系统正常，故有

$$A = A_1 A_2 \cdots A_n,$$

这里 A_k 是第 k 个元件的工作系数. 以 μ_{k_1} 及 μ_{k_2} 分别表第 k 个元件的平均寿命及平均更新时间, 则由 §8.2(23) 得

$$A=\prod_{k=1}^{n}\frac{\mu_{k_1}}{\mu_{k_1}+\mu_{k_2}}.$$

例 3 问题完全与例 2 相同, 只是由于考虑更新系统, 可靠性指标稍有改变: 把 (10) 及 (11) 中的 $q_j(t)$ 换为 $\frac{\mu_{j_2}}{\mu_{j_1}+\mu_{j_2}}$. 于是近似解中的 $q_j(t)$ 也只需这样改变即可.

关于可靠性理论, 我们只叙述至此为止. 由于实际的需要, 这方面的研究很多, 发展很快, 文献也很丰富, 但都偏重于实用. 从数学观点看来, 比较严格的叙述见 [44],[45].

习 题 解 答

第 一 章

7. $\frac{2}{4!}=\frac{1}{12}$

8. $\binom{T-1}{k-1}s^k r^{T-k}/(s+r)^T$

9. $n!/n^n$

10. $\frac{(N-1)^n}{N^n}$ $(n \geqslant N)$

11. $\frac{1}{2}$

12. 注意“四位偶数”必须是四位有效数字，即不能以 0 开头.

$\frac{41}{90}$

13. (i) $\binom{N}{n}n!/N^n$.

(ii) 解. 每次都可以从 N 个数中取出一个，故取 n 次所有可能的取法有 N^n 种，取出来的有三类： $<M$, $=M$, $>M$.

如果我们固定 k_1 次是取到 $<M$ 的数，固定 k_2 次是取到 $>M$ 的数，当然其余的一定是取到 M 的.

当次数固定后， $<M$ 的有 $(M-1)^{k_1}$ 种可能的取法 (因为每一次都可以从 $M-1$ 个数中取一个)， $>M$ 的有 $(N-M)^{k_2}$ 种可能的取法，而 $=M$ 的只有一种取法即全是 M, 所以可能的取法有

$$(M-1)^{k_1}(N-M)^{k_2}$$

但是次数可以有不同的固定方式，共有 [1]

$$C_{k_1}^n \times C_{k_2}^{n-k_1} \text{ 种}$$

因此， k_1 次取到 $<M$ 的数， k_2 次取到 $>M$ 的数的可能的取法有

$$C_{k_1}^n C_{k_2}^{n-k_1}(M-1)^{k_1}(N-M)^{k_2} \text{ 种}.$$

设 A 表事件“从小到大第 m 个数等于 M”.

事件 A 出现，也就是 k_1 次取到 $<M$ 的数， k_2 次取到 $>M$ 的数， $0 \leqslant k_1 \leqslant m-1, 0 \leqslant k_2 \leqslant n-m$. 因此 A 包含的所有可能的取法有

$$\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{n-m} C_{k_1}^n C_{k_2}^{n-k_1}(M-1)^{k_1}(N-M)^{k_2} \text{ 种}.$$

所以

$$P(A)=\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{n-m} C_{k_1}^n C_{k_2}^{n-k_1}(M-1)^{k_1}(N-M)^{k_2}/N^n.$$

14. 这题在统计物理中有用，所以我们给出两种详细的解法.

解一. 首先考虑基本事件的总数.

[1] 简写 $\binom{n}{k}$ 为 C_k^n, 以后同此

匣可辨，故可将匣子按一定顺序排列起来，将 n 个球投入这排列起来的 N 个匣子，我们关心的只是每一个匣中的球数而不管它是哪些球，因为球是不可辨的.

每一个基本事件 $\omega=(x_1,x_2,\cdots,x_N)$, x_j 表示第 j 匣中的球数， $0\le x_j\leqslant n$.

$$x_1+x_2+\cdots+x_N=n$$

因此，基本事件的总数为 $(1+x+x^2+\cdots+x^n)^N$ 的展开式中 x^n 的系数

$$(1+x+x^2+\cdots+x^n)^N=\left(\frac{1-x^{n+1}}{1+x}\right)^N=\left(\sum_{l=0}^{N}(-1)^lC_l^N x^{(n+1)l}\right)\left(\sum_{k=0}^{\infty}C_k^{N+k-1}x^k\right).$$

所以 x^n 的系数应为 $C_n^{N+n-1}=C_{N-1}^{N+n-1}$.

(i) 没有空匣所包含的基本事件， $0<x_j\leqslant n$, 即为 $(x+x^2+\cdots+x^n)^N$ 展开式中 x^n 的系数，或 $(1+x+x^2+\cdots+x^{n-1})^N$ 展开式中 x^{n-N} 的系数

$$(1+x+x^2+\cdots+x^{n-1})^N=\left(\frac{1-x^n}{1-x}\right)^N=\left(\sum_{l=0}^{N}(-1)^lC_l^N x^{nl}\right)\left(\sum_{k=0}^{\infty}C_k^{N+k-1}x^k\right).$$

所以 x^{n-N} 的系数应为 $C_{n-N}^{N+(n-N)-1}=C_{n-N}^{n-1}=C_{N-1}^{n-1}$. 若令 A_1"没有空匣"

$$P(A_1)=\frac{C_{N-1}^{n-1}}{C_{N-1}^{N+n-1}}.$$

(ii) 固定 m 个空匣，即将 n 个球投入 $N-m$ 个匣中，按照 (i) 应为

$$\begin{aligned}(x+x^2+\cdots+x^n)^{N-m}&=x^{N-m}\left(\frac{1-x^n}{1-x}\right)^{N-m}\\&=x^{N-m}\left(\sum_{l=0}^{N-m}(-1)^lC_l^{N-m}x^{nl}\right)\left(\sum_{k=0}^{\infty}C_k^{N-m+k-1}x^k\right).\end{aligned}$$

x^n 的系数应为 $C_{n-N+m}^{n-1}=C_{N-m-1}^{n-1}$.

但是，由于匣可辨， m 个空匣有 C_m^N 种不同取法，令 A_m"恰好有 m 个空匣"

$$P(A_m)=\frac{C_m^NC_{N-m-1}^{n-1}}{C_{N-1}^{N+n-1}}.$$

(iii) 指定的 m 个匣中正好有 j 个球，而其余的 $N-m$ 个匣中正好有 $n-j$ 个球，按照本题开始时的分析，前者有 C_j^{m+j-1} 种投法，后者有 $C_{N-m-1}^{N-m+n-j-1}$ 种投法，故所求的概率为

$$\frac{C_j^{m+j-1}C_{N-m-1}^{N+n-m-j-1}}{C_{N-1}^{N+n-1}}$$

解二. 现在我们用另一个更为直观的方法来分析这个问题，同样将匣子按照一定的顺序排列起来，则 n 个球的每一种投法可以表示如下

$$|*\ *|*|*\ *\ *|\cdots|*|*\ *|.$$

每一个 "|" 表示匣子的壁， "*" 表示一个球. 如果我们将每一个壁和每一个球都看成是一个位置，当然，最外二层必须是两个壁，而 n 个球每一种投法就对应于中间的 $N-1$ 壁和 n 个球不同的位置 (球之间

和壁之间的次序不用考虑)，也就是说 $N+n-1$ 个位置被壁占据了 $N-1$ 个. 因此总的投法有 C_{N-1}^{N+n-1} 种.

(i) 没有空匣，即每一个壁必须在两个球之间，亦即要从 n 个球所形成的空隙中间插入 $N-1$ 个壁，共有 C_{N-1}^{n-1} 种.

$$P(A_1)=\frac{C_{N-1}^{n-1}}{C_{N-1}^{N+n-1}}.$$

(ii) 有 m 个空匣，相当于少了 m 个壁，对固定的 m 个空匣来说，相当于在 $n-1$ 个空隙中安插入 $N-m-1$ 个壁，共有 C_{N-m-1}^{n-1} 种. m 个空匣，有 C_m^N 种固定法. 故

$$P(A_m)=\frac{C_m^N C_{N-m-1}^{n-1}}{C_{N-1}^{N+n-1}}.$$

(iii) 指定的 m 个匣中有 j 球，共有 C_j^{m+j-1} 种不同的放法，其余 $N-m$ 个匣中有 $n-j$ 个球共有 $C_{N-m-1}^{N+n-m-j-1}$ 种不同的放法.

故所求的概率为

$$\frac{C_j^{m+j-1}C_{N-m-1}^{N+n-m-j-1}}{C_{N-1}^{N+n-1}}.$$

15. $\frac{m^k-(m-1)^k}{n^k}$ (参考 §1.5 例 4).

16. (ii) $m\cdot n^k/(m+n)^{k+1}$.

17. 设 x,y 为二船到达时刻，一切可能值为 $0\leqslant x\leqslant 24$, $0\leqslant y\leqslant 24$.

所求值满足 $y-x\leqslant 1$, $x-y\leqslant 2$.

所求概率为 0.121.

18. $\left(1-\frac{t_0}{T_2-T_1}\right)^2$.

注意任何一个等待的时间都限制在 T_1 至 T_2 之间.

19. (i) 试验的结果已知，要求事前概率，故利用巴叶斯公式

$$P(U_i|v)=\frac{P(v|U_i)P(U_i)}{\sum\limits_{j=0}^{N}P(v|U_j)P(U_j)}.$$

若 $i<v$ 显然 $P(U_i|v)=0$, $i\geqslant v$

$$P(U_i|v)=\frac{\frac{C_v^iC_{n-v}^{N-i}}{C_n^N}}{\sum_{j=v}^{N}\frac{C_v^jC_{n-v}^{N-j}}{C_n^N}}=\frac{C_v^iC_{n-v}^{N-i}}{\sum_{j=v}^{N}C_v^jC_{n-v}^{N-j}}.$$

(ii) $\frac{\sum_{i=v}^{N}\frac{i-v}{N-n}C_v^iC_{n-v}^{N-i}}{\sum_{i=v}^{N}C_v^iC_{n-v}^{N-i}}$.

20. $\frac{7}{18}$(利用全概率公式).

21. (i) $\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}$(利用乘法公式).

(ii) $C_{n_1}^{n}\prod\limits_{i=0}^{n_1-1}(b+ic)\prod\limits_{j=0}^{n_2-1}(r+jc)\Big/\prod\limits_{k=0}^{n-1}(b+r+kc)$.

22. 用归纳法，设第 n 次取得黑球的概率为 p_n，显然 $p_1=\frac{b}{b+r}$. 假定 $p_n=\frac{b}{b+r}$，则第一次得黑 (红) 球，第 $n+1$ 次得黑球的概率为 $\frac{b+c}{b+r+c}\left(\frac{b}{b+r+c}\right)$，故

$$p_{n+1}=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}+\frac{r}{b+r}\cdot\frac{b}{b+r+c}=\frac{b}{b+r}.$$

23. 证. 用数学归纳法.

首先，证当 $m=1$, 对一切 n 成立. 令 B_j 表示 "第 j 次取到黑球",R_j 表示 "第 j 次取到红球" $P(B_1B_n)=P(B_1)P(B_n|B_1)$. 而 $P(B_1)=\frac{b}{b+r}$, 求 $P(B_n|B_1)$ 时，可把它看作从 $b+c$ 个黑球和 r 个红球开始第 $n-1$ 次得黑球的概率，由上知

$$P(B_n|B_1)=\frac{b+c}{b+r+c}.$$

故

$$P(B_1B_n)=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}.$$

其次，假定对 $m-1$ 以及一切 n 成立，来证对 m 以及一切 n 成立.

$$P(B_mB_n)=P(B_mB_n|B_1)P(B_1)+P(B_mB_n|R_1)P(R_1),$$

其中 $P(B_mB_n|B_1)$ 可以看作是从 $b+c$ 个黑球和 r 个红球中第 $m-1$ 次和 $n-1$ 次都得黑球的概率，由归纳法假定知

$$P(B_mB_n|B_1)=\frac{b+c}{b+r+c}\cdot\frac{b+2c}{b+r+2c}.$$

又 $P(B_mB_n|R_1)$ 可以看作是从 b 个黑球和 $r+c$ 个红球中第 $m-1$ 次和 $n-1$ 次都得黑球的概率，由归纳法假定知

$$P(B_mB_n|R_1)=\frac{b}{b+r+c}\cdot\frac{b+c}{b+r+2c}.$$

于是

$$P(B_mB_n)=\frac{b+c}{b+r+c}\cdot\frac{b+2c}{b+r+2c}\cdot\frac{b}{b+r}+\frac{b}{b+r+c}\cdot\frac{b+c}{b+r+2c}\cdot\frac{r}{b+r}=\frac{b(b+c)}{(b+r)(b+r+c)}.$$

于是 (i) 得证. (ii) 之证类似.

24. 利用全概率公式.

25. 甲 (乙) 胜的概率为 $\sum\limits_{k=0}^{\infty}(1-p_1)^k(1-p_2)^kp_1\Big(\sum\limits_{k=0}^{\infty}(1-p_1)^{k+1}(1-p_2)^kp_2\Big)$.

26. $\binom{a}{1}\binom{a+b-1}{k}\Big/\binom{a+b-k}{1}\binom{a+b}{k}=\frac{a}{a+b}$.

27. 所求概率为 $\frac{a}{a+b}$. 袋中只剩白球时取出的球必为 b 个黑球， j 个白球 ($j=0,1,\cdots,a-1$), 计算出每一种情形的概率 p_j, 再相加.

28. 提示：设袋中有 A 个球， a 个是白的，不还原随机取出，第 k 次才取得白球的概率为

$$p_k=\frac{a(A-a)(A-a-1)\cdots(A-a-k+2)}{A(A-1)(A-2)\cdots(A-k+1)}.$$

由 $1=\sum\limits_{k=1}^{\infty}p_k=\sum\limits_{k=1}^{A-a+1}p_k$, 即得所要证的等式.

29. (i) 0.612, (ii) 0.997.

30. $\sum\limits_{k=1}^{n}\frac{(-1)^{k-1}}{k!}$ (参看 §1.4 例 7).

31. 解. 令 A_i: "第一次从坛 I 取出 i 个白球", B_i: "从坛 II 中取出 i 个白球", C: "从坛 III 中取到一个白球".

现在是已知试验结果 (从坛 III 中取到一个白球), 求 A_5 的事后概率，所以利用巴叶斯公式

$$P(A_5|C)=\frac{P(C|A_5)P(A_5)}{\sum\limits_{i=0}^{5}P(C|A_i)P(A_i)}.$$

因此只需求出 $P(C|A_i),\ i=0,1,\cdots,5$.

但是从坛 I 到坛 III 中间经过坛 II, 所以 $P(C|A_i)$ 与从坛 II 取出的球有关， C 必与 B_j 之一同时出现，故利用全概率公式

$$P(C|A_i)=\sum_j P(C|B_jA_i)P(B_j|A_i).$$

下面就来分别求出它们.

$$\begin{aligned}
&P(C|A_0)=0,\\
&P(C|A_1)=P(C|B_1A_1)P(B_1|A_1)=\frac{1}{3}\cdot\frac{C_2^4C_1^1}{C_3^5}=\frac{1}{5}\\
&P(C|A_2)=\frac{1}{3}\cdot\frac{C_1^2C_2^3}{C_3^5}+\frac{2}{3}\cdot\frac{C_2^2C_1^3}{C_3^5}=\frac{2}{5},\\
&P(C|A_3)=\frac{1}{3}\cdot\frac{C_1^3C_2^2}{C_3^5}+\frac{2}{3}\cdot\frac{C_2^3C_1^2}{C_3^5}+\frac{C_3^3}{C_3^5}=\frac{3}{5},\\
&P(C|A_4)=\frac{2}{3}\cdot\frac{C_2^4C_1^1}{C_3^5}+\frac{C_3^4}{C_3^5}=\frac{4}{5},\\
&P(C|A_5)=1.
\end{aligned}$$

另一方面

$$\begin{aligned}
&P(A_1)=\frac{C_4^5C_1^5}{C_5^{10}}=\frac{25}{252},\ P(A_2)=P(A_3)=\frac{100}{252},\\
&P(A_4)=\frac{25}{252},\quad P(A_5)=P(A_0)=\frac{1}{252}.
\end{aligned}$$

于是

$$P(A_5|C)=\frac{\frac{1}{252}}{\frac{1}{252}+\frac{5}{252}+\frac{40}{252}+\frac{60}{252}+\frac{20}{252}}=\frac{1}{126}.$$

32. 甲 (乙) 获胜的概率是 $\frac{\alpha^2}{1-2\alpha\beta}\left(\frac{\beta^2}{1-2\alpha\beta}\right)$

找出甲 (乙) 获胜的所有可能的情形，然后用乘法公式求得每种情形的概率.

本题用差分方程法亦可求解.

第　二　章

2. (i)

$$F_\eta(x)=\begin{cases}1-F\left(\dfrac{1}{x}\right)+F(0), & x>0,\\ F(0), & x=0,\\ F(0)-F\left(\dfrac{1}{x}\right), & x<0;\end{cases}$$

密度为

$$f_\eta(x)=\begin{cases}\dfrac{1}{x^2}p\left(\dfrac{1}{x}\right), & x\neq 0,\\ p(0), & x=0.\end{cases}$$

(ii)

$$F_\eta(x)=\begin{cases}F(x)-F(-x), & x\geqslant 0,\\ 0, & x<0;\end{cases}$$

$$f_\eta(x)=\begin{cases}p(x)+p(-x), & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

3. (i) 因 $F(x)$ 连续，故 $\mathcal{B}_1$- 可测，从而 η 是随机变数；

(ii) $F_\eta(x)=\begin{cases}0, & x<0,\\ x, & 0\leqslant x<1,\\ 1, & x\geqslant 1,\end{cases}$ 即 $[0,1]$ 上的均匀分布.

4. (i) $a=\frac{1}{2}$.

6. 解. $x\int_x^\infty \frac{1}{z}\,dF(z)\leqslant\int_x^\infty dF(z)\to 0,\ x\to\infty$.

即 $$\lim_{x\to\infty} x\int_x^\infty \tfrac{1}{z}\,dF(z)=0.$$

类似可证 $$\lim_{x\to-\infty} x\int_{-\infty}^x \tfrac{1}{z}\,dF(z)=0.$$

今证： $$\lim_{x\to+0} x\int_x^\infty \tfrac{1}{z}\,dF(z)=0.$$

任给 $\alpha>\varepsilon>0$, 存在一个 $x_0>0$, 使 $\int_{+0}^{x_0}dF(z)<\frac{\varepsilon}{2}$.

对于取定的 x_0, 只要 $0<x<\frac{x_0\varepsilon}{2}$, 就有

$$x\int_{x_0}^\infty\frac{1}{z}\,dF(z)\leqslant\frac{x}{x_0}\int_{x_0}^\infty dF(z)\leqslant\frac{x}{x_0}<\frac{\varepsilon}{2},$$

$$x\int_x^\infty\frac{1}{z}\,dF(z)=x\int_{x_0}^\infty\frac{1}{z}\,dF(z)+x\int_{+0}^{x_0}\frac{1}{z}\,dF(z),$$

$$\leq\frac{x}{x_0}\int_{x_0}^\infty dF(x)+\int_{+0}^{x_0}dF(z)<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

故得证. 类似有

$$\lim_{x\to\infty} x\int_{-\infty}^x\frac{1}{z}\,dF(x)=0.$$

7. $p_{\xi_1}(x_1)=\frac{1}{\Gamma(k_1)}x_1^{k_1-1}e^{-x_1},\ 0<x_1<\infty$.

$p_{\xi_2}(x_2)=\frac{1}{\Gamma(k_1+k_2)}x_2^{k_1+k_2-1}e^{-x_2},\ 0<x_2<\infty$.

8. 解. 书上有 500 个错字，可以看作是将这 500 个错字投入这本书中去，每投一个字看作是一次试验，这个试验有二个可能的结果，落到所指定的一页上，或没有落在指定的一页上，前者的概率是 $\frac{1}{500}$，后者的概率是 $\frac{499}{500}$. 500 个错字可看作是做了 500 次这样的试验，由上面的分析可知是一个贝努里试验，因此给定的一页上至少有三个错字的概率是

$$\sum_{k=3}^{500}C_k^{500}\left(\frac{1}{500}\right)^k\left(\frac{499}{500}\right)^{500-k}=\left(\frac{499}{500}\right)^{500}\sum_{k=3}^{500}\frac{C_k^{500}}{499^k}.$$

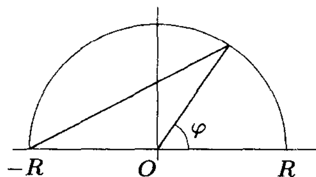
9. 解. 设弦长是 x

$$x = 2R\cos\frac{\varphi}{2}, \quad -\pi \le \varphi \le \pi.$$

而 φ 在 $(-\pi, \pi)$ 上均匀分布.

弦长的分布函数是：当 $0 < x \leqslant 2R$ 时,

$$F(x) = P\Big(2R\cos\frac{\varphi}{2} \leqslant x\Big);$$



此外，又有

$$F(x) = \begin{cases} 0, & x \leqslant 0, \\ 1, & x > 2R. \end{cases}$$

仿照 §2.8 例 2, 得分布密度为

$$f(x) = F'(x) = \begin{cases} 0, & x \leqslant 0, x > 2R, \\ \dfrac{2}{\pi} \cdot \dfrac{1}{\sqrt{4R^2 - x^2}}, & 0 < x \leqslant 2R. \end{cases}$$

10. $F(x) = \begin{cases} 0, & x < 0, \\ x(2a-x)/a^2, & 0 \leqslant x \leqslant a, \\ 1, & x > a. \end{cases}$

11. 解. 设所抛的 n 个点的横坐标是 $\xi_1, \cdots, \xi_n$, 将其按大小顺序排起来，得到 $\xi_{(1)} \leqslant \xi_{(2)} \le \cdots \leqslant \xi_{(n)}$.

(i) 求 $\xi_{(k)}$ 的分布密度.

$$P(\xi_{(k)} \leqslant x) = \sum_{j=k}^{n} \Sigma' P(\xi_{i_1} \leqslant x, \cdots, \xi_{i_j} \leqslant x;\ \xi_{i_{j+1}} > x, \cdots, \xi_{i_n} > x),$$

其中 Σ' 是对 C_j^n 组取法 $(i_1, \cdots, i_j, i_{j+1}, \cdots, i_n)$ 求和.

由于 $(\xi_1, \cdots, \xi_n)$ 的独立性,

$$\begin{aligned} &P(\xi_{i_1} \leqslant x, \cdots, \xi_{i_j} \leqslant x;\ \xi_{i_{j+1}} > x, \cdots, \xi_{i_n} > x) \\ &= P(\xi_{i_1} \leqslant x) \cdots P(\xi_{i_j} \leqslant x) P(\xi_{i_{j+1}} > x) \cdots P(\xi_{i_n} > x) = [F(x)]^j [1 - F(x)]^{n-j}. \end{aligned}$$

于是

$$P(\xi_{(k)} \leqslant x) = \sum_{j=k}^{n} C_j^n [F(x)]^j [1 - F(x)]^{n-j} = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} y^{k-1}(1-y)^{n-k}\, dy.$$

$\xi_{(k)}$ 的分布密度

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} F'(x), \quad 0 < x < a.$$

以 $F(x) = \frac{x}{a}$, $F'(x) = \frac{1}{a}$ 代入得

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} \frac{x^{k-1}(a-x)^{n-k}}{a^n}, \quad 0 < x < a.$$

(ii) 求 $\xi_{(k)}$ 和 $\xi_{(m)}$ 的联合分布 $(k < m)$. 对 $y < x$,

$$\begin{aligned} P(\xi_{(k)} \leqslant x, \xi_{(m)} \leqslant y) = P(\xi_{(m)} \leqslant y) &= \sum_{j=m}^{n} C_j^n [F(y)]^j [1 - F(y)]^{n-j} \\ &= \frac{n!}{(m-1)!(n-m)!} \int_0^{F(y)} y^{m-1}(1-y)^{n-m}\, dy. \end{aligned}$$

对 $y > x$,

$$P(\xi_{(k)} \leqslant x,\ \xi_{(m)} \leqslant y) = \sum_{l=m}^{n}\sum_{j=k}^{l}\Sigma' P(\xi_{i_1} \leqslant x, \cdots, \xi_{i_j} \leqslant x; x < \xi_{i_{j+1}} \leqslant y, \cdots, x < \xi_{i_l} \leqslant y,\ \xi_{i_{l+1}} > y, \cdots, \xi_{i_n} > y),$$

其中 Σ' 是对所有可能的分组 $(i_1, \cdots, i_j; i_{j+1}, \cdots, i_l; i_{l+1}, \cdots, i_n)$ 求和. 同样由 $(\xi_1, \cdots, \xi_n)$ 的独立性得

$$P(\xi_{(k)} \leqslant x,\ \xi_{(m)} \leqslant y) = \sum_{l=m}^{n}\sum_{j=k}^{l} C_l^n C_j^l [F(x)]^j [F(y) - F(x)]^{l-j} [1 - F(y)]^{n-l}$$

$$= \frac{n!}{(k-1)!(m-k-1)!(n-m)!}\int_0^{F(x)} dy_1 \int_0^{F(y)} y_1^{k-1}(y_2 - y_1)^{m-k-1}(1-y_2)^{n-m}\,dy_2,$$

于是 $\xi_{(k)}, \xi_{(m)}$ 的联合分布密度是

$$p(x,y) = \begin{cases} 0, \quad 0 < y < x < a;, \\ \dfrac{n!}{(k-1)!(m-k-1)!(n-m)!}[F(x)]^{k-1}[F(y) - F(x)]^{m-k-1} \\ \quad \times [1 - F(y)]^{n-m} F'(x) F'(y), \quad 0 < x < y < a. \end{cases}$$

将 $F(x) = \frac{x}{a}, F(y) = \frac{y}{a}, F'(x) = F'(y) = \frac{1}{a}$ 代入，得

$$p(x,y) = \begin{cases} 0, \quad 0 < y < x < a; \\ \dfrac{n!}{(k-1)!(m-k-1)!(n-m)!}\dfrac{x^{k-1}(y-x)^{m-k-1}(a-y)^{n-m}}{a^n}, \quad 0 < x < y < a. \end{cases}$$

12. 直接利用上题的结果令 $k = 1, m = n$, 即可.

13. $f_{\xi\times\eta}(x) = \frac{1}{\pi}\int_0^\infty e^{-\frac{z^4+x^2}{2z^2}}\frac{1}{z}\,dz.$

14. (i) $P_\zeta(x) = \begin{cases} \dfrac{1}{(x+1)^2}, & x > 0, \\ 0, & x \leqslant 0 \end{cases}$; (ii) $P_\zeta(x) = \begin{cases} \dfrac{1}{2x^2}, & x > 1, \\ 0, & x \leqslant 0, \\ \dfrac{1}{2}, & 0 < x \leqslant 1. \end{cases}$

15. (i) $\frac{2}{\pi(x^2+4)}$.

(ii) $f(x) = \begin{cases} \dfrac{x+4}{24}, & -4 \leqslant x \leqslant 0, \\ \dfrac{1}{6}, & 0 \leqslant x \leqslant 2, \\ \dfrac{6-x}{24}, & 2 \leqslant x \leqslant 6, \\ 0, & x > 6\text{或}x < -4. \end{cases}$

(iii) $\frac{1}{4\alpha}e^{-\frac{|x|}{\alpha}}\left(1 + \frac{|x|}{\alpha}\right)$.

16. 证. (ξ, η) 的密度函数是

$$f(x,y) = Cx^{\frac{m}{2}-1}y^{\frac{n}{2}-1}e^{-\frac{x+y}{2}},\ (x \geqslant 0, y \geqslant 0),$$

其中
$$C = \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}.$$

设 $U = \xi + \eta$ 与 $V = \xi/\eta$ 的联合密度是 $g(u,v)$, 令 $u = x + y$, $v = x/y$, 则

$$g(u,v) = f(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|; \quad \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \frac{u}{(1+v)^2}$$

故 $$g(u,v)=Cu^{\frac{m+n}{2}-1}e^{-u/2}\cdot v^{\frac{m}{2}-1}(1+v)^{-\frac{m+n}{2}},$$
即 $g(u,v)$ 是 u 的函数及 v 的函数的积，故知 U,V 独立.

17. 证. 由独立性知 ξ,η 的联合密度为 $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$.

又 $U=\xi^2+\eta^2$, $V=\xi/\eta$ 有两组解为

$$\xi_1=V\sqrt{U}/W,\ \eta_1=\sqrt{U}/W;\ \xi_2=-\xi_1,\ \eta_2=-\eta_1,$$

其中 $W=\sqrt{1+V^2}$. 对第一组解

$$J_1=\frac{\partial(\xi_1,\eta_1)}{\partial(U,V)}=\begin{vmatrix}\frac{V}{2W\sqrt{U}} & \frac{\sqrt{U}}{W^3}\\ \frac{1}{2W\sqrt{U}} & \frac{-V\sqrt{U}}{W^3}\end{vmatrix}=-\frac{1}{2(1+V^2)},$$

故 $|J_1|=|J_2|=\frac{1}{2(1+V^2)}$. 根据 §2.8 定理一，注 1 知 (U,V) 的联合密度 $g(u,v)$ 如下式所示.

因 $g(u,v)=(\pi)^{-1}e^{-\frac{u}{2}}/2(1+v^2)$ 可表为 u 的函数与 v 的函数的积， $u\geqslant 0$, v 任意，从而 U,V 独立.

18. 令 $U=\xi+\eta,V=\eta$, 则 $U\cdot V$ 的密度为

$$g(u,v)=(\ln\theta)^2\theta^{-u},\ 0<v<u<\infty.$$

而 U 的分布密度是 $(\ln\theta)^2u\theta^{-u}$, $0<u<\infty$. V 的分布密度是 $(\ln\theta)\theta^{-v}$, $0<v<\infty$. 故 U,V 不独立.

19. 解. ξ,η 的联合分布密度是

$$f(x,y)=p(x)p(y)=\begin{cases}e^{-(x+y)}, & 0<x,y<\infty,\\ 0, & \text{其他}.\end{cases}$$

令 $u=\xi+\eta,\xi=uv$, $|J|=\begin{vmatrix}v & u\\ 1-v & -u\end{vmatrix}=-u,$

$$v=\frac{\xi}{\xi+\eta},\quad \eta=u(1-v),$$

所以 u,v 的联合分布是

$$g(u,v)=ue^{-u},\ 0<u<\infty,\ 0<v<1.$$

u 的分布密度是 $h(u)=\int_0^1 g(u,v)\,dv=ue^{-u}, 0<u<\infty.$

v 的分布密度是 $\widetilde{h}(v)=\int_0^\infty ue^{-u}\,du=1,\ 0<v<1.$

由 $g(u,v)=h(u)\widetilde{h}(v)$ 成立，所以 u,v 独立.

20. 提示：利用概率想法，注意 $b(k,n,p)$ 是 n 次试验中恰好成功 k 次的概率，而 $B(k,n,p)$ 是 n 次中至多成功 k 次的概率. 利用前式以证明后式.

21. 提示：将等式 $\int_0^p(p-x)^\alpha x^{n-k-1}\,dx=p^{n-k+\alpha}\int_0^1(1-x)^\alpha x^{n-k-1}\,dx$ 的右方分部积分，得

$$\int_0^p(p-x)^\alpha x^{n-k-1}\,dx=p^{n-k+\alpha}\frac{\alpha!(n-k-1)!}{(\alpha+n-k)!},(n-k>0,\alpha\geqslant 0)\tag{1}$$

将此关系代入所要证明的等式的左方，得

$$\begin{aligned}\sum_{\alpha=0}^{k}C_{k-\alpha}^{n}p^{n-k+\alpha}q^{k-\alpha}&=\sum_{\alpha=0}^{k}\int_0^p(p-x)^\alpha x^{n-k-1}\,dx\frac{n!q^{k-\alpha}}{\alpha!(n-k-1)!(k-\alpha)!}\\&=\sum_{\alpha=0}^{k}\int_0^p(p-x)^\alpha x^{n-k-1}\,dx(1-p)^{k-\alpha}\frac{k!}{\alpha!(k-\alpha)!}\Big/\frac{k!(n-k-1)!}{n!}\\&=\int_0^p(1-x)^k x^{n-k-1}dx\Big/\frac{k!(n-k-1)!}{n!}.\end{aligned}$$

其次，在 (1) 中取 $p=1, \alpha=k$, 则 $\frac{k!(n-k-1)!}{n!}=\int_0^1(1-x)^k x^{n-k-1}\,dx$, 以此式代入上式分母，即得所欲证.

22. 提示：注意到 $q=\alpha\beta$, $p=-(\alpha+\beta)$, 利用乘积以及和的分布，即得 p,q 的分布密度为

$$f_p(x)=\begin{cases}\dfrac{2-|x|}{4}, & |x|\leqslant 2,\\ 0, & |x|>2.\end{cases}\quad f_q(x)=\begin{cases}-\dfrac{1}{2}\log x, & 0<x\leqslant 1,\\ -\dfrac{1}{2}\log(-x), & -1\leqslant x<0,\\ 0, & \text{其他}.\end{cases}$$

23. $E\xi=a,\ D\xi=a(1+\alpha)$.

24. $c=2h^2$, $E\xi=\frac{\sqrt{\pi}}{2|h|}$, $D\xi=\frac{4-\pi}{4h^2}$.

25. (i) 0, $\frac{\pi^2}{12}-\frac{1}{2}$; (ii) 0, $\frac{1}{\alpha+3}$.

26. 提示：参考 §2.9(16).

29. 提示：利用 Cauchy 不等式.

$$\left(\int f\cdot g\,dF\right)^2\leq\int f^2\,dF\cdot\int g^2\,dF.$$

30. 提示：先算

$$q=\frac{1}{2\pi\sqrt{1-r^2}}\iint_{(x-a_1)(y-a_2)<0}e^{-\frac{1}{2(1-r^2)}[(x-a_1)^2-2r(x-a_1)(y-a_2)+(y-a_2)^2]}dxdy$$

31. 证. (ξ,η) 的联合分布密度是

$$f(x,y)=\frac{1}{2\pi\sqrt{1-R^2}}e^{-\frac{1}{2(1-R^2)}[x^2-2Rxy+y^2]}$$

$$\begin{aligned}E\max(\xi,\eta)=&\frac{1}{2\pi\sqrt{1-R^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\max(x,y)e^{-\frac{1}{2(1-R^2)}(x^2-2Rxy+y^2)}dxdy\\
=&\frac{1}{2\pi\sqrt{1-R^2}}\left[\int_{-\infty}^{\infty}x\,dx\int_{-\infty}^{x}e^{-\frac{1}{2(1-R^2)}(x^2-2Rxy+y^2)}dy\right.\\
&\left.+\int_{-\infty}^{\infty}ydy\int_{-\infty}^{y}e^{-\frac{1}{2(1-R^2)}(x^2-2Rxy+y^2)}dx\right]\\
=&\frac{1}{\pi\sqrt{1-R^2}}\int_{-\infty}^{\infty}x\,dx\int_{-\infty}^{x}e^{-\frac{1}{2(1-R^2)}(x^2-2Rxy+y^2)}dy\\
=&\frac{1}{\pi\sqrt{1-R^2}}\int_{-\infty}^{\infty}xe^{-\frac{x^2}{2}}dx\int_{-\infty}^{x}e^{-\frac{(y-Rx)^2}{2(1-R^2)}}dy\\
=&\frac{1}{\pi}\int_{-\infty}^{\infty}xe^{-\frac{x^2}{2}}dx\int_{-\infty}^{\sqrt{\frac{1-R}{1+R}}x}e^{-\frac{y^2}{2}}dy\\
=&\frac{1}{\pi}\left[-e^{-\frac{x^2}{2}}\int_{-\infty}^{\sqrt{\frac{1-R}{1+R}}x}e^{-\frac{y^2}{2}}dy\right]_{-\infty}^{\infty}+\frac{1}{\pi}\sqrt{\frac{1-R}{1+R}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(1+\frac{1-R}{1+R})x^2}dx\\
=&\frac{1}{\pi}\sqrt{\frac{1-R}{1+R}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{1+R}}dx=\frac{1}{\pi}\sqrt{\frac{1-R}{1+R}}\cdot\sqrt{2\pi}\cdot\sqrt{\frac{1+R}{2}}=\sqrt{\frac{1-R}{\pi}}.\end{aligned}$$

32. 解. $r_{\xi\eta}=\frac{E\xi\eta-E\xi\cdot E\eta}{\sqrt{D\xi}\cdot\sqrt{D\eta}}$,

$$E\xi=E(\alpha X+\beta Y)=0,$$

$$E\eta=E(\alpha X-\beta Y)=0,$$

$$E\xi\eta=E(\alpha X+\beta Y)(\alpha X-\beta Y)=E(\alpha^2X^2-\beta^2Y^2)=(\alpha^2-\beta^2)\sigma^2,$$

$$D\xi=E(\alpha X+\beta Y)^2=(\alpha^2+\beta^2)\sigma^2,$$

$$D\eta=E(\alpha X-\beta Y)=(\alpha^2+\beta^2)\sigma^2,$$

所以 $r_{\xi\eta}=\frac{\alpha^2-\beta^2}{\alpha^2+\beta^2}$.

33. 提示：令 $\xi=\frac{X_1-a}{\sigma}$, $\eta\frac{X_2-a}{\sigma}$ 得 $\max(X_1,X_2)=a+\sigma\max(\xi,\eta)$, 然后利用 31 题.

34. 证： $P(|X|\geqslant x)=P(f(|X|)\geqslant f(x))\leqslant[f(x)]^{-1}\int_{|y|\geqslant x}f(y)\,dF(y)\leqslant E(f(|X|))/f(x)$.

35. 提示：利用上题.

36. (i) $f_\eta(x)=\frac{1}{x}f_\xi(-\log x)$; $0<x<1$.

(ii)

$$F_\eta(x)=P(\operatorname{tg}\xi\leqslant x)=\sum_{n=-\infty}^{\infty}P\Big(\operatorname{tg}\xi\leq x,-\frac{\pi}{2}+n\pi<\xi\leqslant\frac{\pi}{2}+n\pi\Big)$$

$$=\sum_{n=-\infty}^{\infty}P\Big(-\frac{\pi}{2}+n\pi<\xi\leqslant\operatorname{arctg}x+n\pi\Big)=\sum_{n=-\infty}^{\infty}\int_{-\frac{\pi}{2}+n\pi}^{\operatorname{arctg}x+n\pi}f_\xi(u)\,du,$$

$$f_\eta(x)=\frac{1}{1+x^2}\sum_{n=-\infty}^{\infty}f_\xi(\operatorname{arctg}x+n\pi).$$

(iii) $f_\eta(x)=f_\xi(tg\,x)/\cos^2x$.

37. 提示：前式显然；后式由 §2.9 定理 1 (iv) 得

$$D\xi\leqslant E\Big(\xi-\frac{b+a}{2}\Big)^2\leq E\Big(\frac{b+a}{2}-a\Big)^2=\Big(\frac{b-a}{2}\Big)^2.$$

41. 证. 设 x_i 的分布是 F.

$$P\Big(\sum_{i=1}^{k}x_i^2\geqslant\lambda_k\Big)=\underset{\sum x_j^2\geqslant\lambda_k}{\int\cdots\int}dF(x_1)\cdots dF(x_k)$$

$$\leqslant\frac{1}{\lambda_k}\underset{\sum x_j^2\geqslant\lambda_k}{\int\cdots\int}\Big(\Sigma x_i^2\Big)dF(x_1)\cdots dF(x_k)$$

$$\leqslant\frac{1}{\lambda_k}\int\cdots\int\Big(\sum x_i^2\Big)dF(x_1)\cdots dF(x_k)=\frac{1}{\lambda_k}\sum_i\int x_i^2\,dF(x_i)=\frac{1}{\lambda}.$$

或者直接利用马尔科夫不等式证明亦可.

42. 提示：半不变量是矩的多项式. 而奇数阶不变量每一项都含有奇数阶矩.

43. $f(t)=pe^{ita}+(1-p)e^{itb}$.

44. $f(t)=\sin at/at$.

45. $f(t)=a^2/(a^2+t^2)$.

46. $f(t)=\frac{p}{1-qe^{it}}$, $E\xi=\frac{q}{p}$, $D\xi=\frac{q}{p}+\frac{q^2}{p^2}$.

49. 解. 各给出两种解法. (i) $\frac{1}{1-it}$ (一) 先试解密度函数

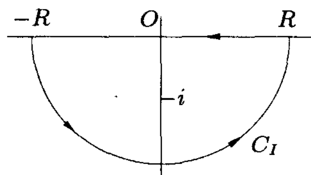
$$g(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}}{1-it}\,dt.$$

考虑复变函数积分

$$\int_{C_I}\frac{e^{-ixz}}{1-iz}\,dz.$$

被积函数在 $z=-i$ 有一个极点.

取闭路 C_I 如图，则



$$\int_{C_I}\frac{e^{-ixz}}{1-iz}\,dz=\int_R^{-R}\frac{e^{-ixt}}{1-it}\,dt+\int_{-\pi}^{0}iR\cdot e^{i\theta}\frac{e^{-ixRe^{i\theta}}}{1-iRe^{i\theta}}\,d\theta,$$

$$\lim_{R\to\infty}\int_R^{-R}\frac{e^{-ixt}}{1-it}\,dt=-\int_{-\infty}^{\infty}\frac{e^{-ixt}}{1-it}\,dt.$$

$$\left|Ri\int_{-\pi}^{0}\frac{e^{-ixRe^{i\theta}}}{1-iRe^{i\theta}}e^{i\theta}\,d\theta\right|\leqslant R\int_{-\pi}^{0}\left|\frac{e^{-ixRe^{i\theta}}}{1-iRe^{i\theta}}\right|d\theta\leqslant\frac{R}{R-1}\int_{-n}^{0}e^{xR\sin\theta}\,d\theta$$

所以当 $x>0$ 时，由控制收敛定理得

$$\lim_{R\to\infty}\int_{-\pi}^{r}iRe^{i\theta}\frac{e^{-ixRe^{i\theta}}}{1-iRe^{i\theta}}\,d\theta=0.$$

另一方面，对任何 $|R|>1$

$$\int_{C_I}\frac{e^{ixz}}{1-iz}\,dz=2\pi i\cdot c_{-1}.$$

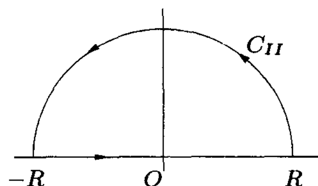
c_{-1} 是被积函数在 $z=-i$ 的残数， $c_{-1}=\frac{e^{-x}}{-i}$.

所以，当 $R\to\infty$, $x\geqslant 0$ 我们得到

$$\int_{-\infty}^{\infty}\frac{e^{-ixt}}{1-it}\,dt=-2\pi i\left(\frac{e^{-x}}{-i}\right)=2\pi e^{-x},$$

即 $$g(x)=e^{-x}.$$

对于 $x<0$, 只须在上半平面取闭路 C_{II} 积分，因为被积函数在上半平面没有奇点，故



$$\int_{C_{II}}\frac{e^{-ixz}}{1-iz}\,dz=0.$$

于是 $g(x)=0$, $x<0$, 这样得到的 $g(x)$ 确实是密度函数， $\int_0^{\infty}e^{-x}\,dx=1$.

(二) 由所给的特征函数容易直接看出

$$\int_0^{\infty}e^{ixt}e^{-x}\,dx=\frac{1}{1-it},\text{亦可知}g(x)=\begin{cases}e^{-x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

(ii) $\frac{1}{(1-it)^n}$.

(一) 先试解 $g(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}}{(1-it)^n}\,dt$.

类似于 (i) 考虑复变函数积分

$$\int_{C_I} \frac{e^{-ixz}}{(1-iz)^n}\,dz.$$

$$\begin{aligned} c_{-1} &= \lim_{z\to -i}\frac{1}{(n-1)!}\frac{d^{n-1}}{dz^{n-1}}\left[(z+i)^n\frac{e^{-ixz}}{(1-iz)^n}\right] \\ &= \lim_{z\to -i}\frac{i^n}{(n-1)!}\cdot\frac{d^{n-1}}{dz^{n-1}}\left[e^{-ixz}\right] \\ &= \frac{i^n}{(n-1)!}(-ix)^{n-1}e^{-x} = \frac{i}{(n-1)!}x^{n-1}e^{-x}. \end{aligned}$$

$$\left|\int_{-\pi}^{0} iRe^{i\theta}\frac{e^{-ixRe^{i\theta}}}{(1-iRe^{i\theta})^n}\,d\theta\right| \longrightarrow 0 \quad x\geqslant 0,\ R\to\infty,$$

故得，对 $x\geqslant 0$

$$g(x)=\frac{x^{n-1}}{(n-1)!}e^{-x}.$$

同样，考虑 $\int_{C_{II}}$, 得 $g(x)=0,\ x<0$, 故

$$g(x)=\begin{cases}\dfrac{x^{n-1}}{(n-1)!}e^{-x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

既然 $\int_0^\infty g(x)\,dx=1$, 于是 $g(x)$ 即为所求.

(二) 如果利用 (i) 的结果

$$\frac{1}{1-it}=\int_0^\infty e^{-x(1-it)}\,dx..$$

两边对 t 求 $n-1$ 次导数，得

$$\frac{(-i)^{n-1}(-1)^{n-1}(n-1)!}{(1-it)^n}=i^{n-1}\int_0^\infty x^{n-1}e^{-x}\cdot e^{itx}\,dx,$$

$$\frac{1}{(1-it)^n}=\int_0^\infty \frac{x^{n-1}e^{-x}}{(n-1)!}\cdot e^{itx}\,dx,$$

故得 $g(x)=\frac{x^{n-1}e^{-x}}{(n-1)!},\ x\geqslant 0$.

50. 提示：为了证明 $f(t)\equiv 1$, 只须证明对应的分布函数是集中于零点的退化分布，或. $E\xi=D\xi=0$. 即 $\int_{-\infty}^{\infty}x^2dF(x)=0$.

考虑 $f(t)$ 在 $t=0$ 的二级差商

$$\begin{aligned} \Delta^2 f(0) &= \frac{1}{4t^2}(f(2t)+f(-2t)-2f(0)) \\ &= \frac{1}{4t^2}\int_{-\infty}^{\infty}(e^{2itx}+e^{-2itx}-2)dF(x) = -\int_{-\infty}^{\infty}\left(\frac{\sin tx}{t}\right)^2 dF(x). \end{aligned}$$

利用法都 (Fatou) 引理，当 $t\to 0$ 时，右端趋于 $-\int_{-\infty}^{\infty}x^2\,dF(x)$, 而左端趋于 0, 故得证 $\int_{-\infty}^{\infty}x^2\,dF(x)=0$.

51. 证.

如果 ξ 的分布函数是 F, 则 $-\xi$ 的分布函数 $\widetilde{F}(x)=1-F(-x-0)$. 设特征函数是实值，即 $f(t)=\overline{f(t)}=f(-t)$.

$$f(t)=Ee^{it\xi},\quad f(-t)=Ee^{-it\xi}.$$

故 ξ 与 $-\xi$ 有相同的特征函数，由唯一性定理知它们有相同的分布函数，即 $F(x)=\widetilde{F}(x)=1-F(-x-0)$, 故分布对称.

反之，若 $F(x)=\widetilde{F}(x)$, 由唯一性定理

$$f(t)=Ee^{it\xi}=Ee^{-it\xi}=f(-t),$$

所以 $f(t)$ 是实的.

52. 证.

考虑 $G(x)=\frac{1}{h}\int_x^{x+h}F(y)\,dy=\frac{1}{h}\int_0^h F(x+y)\,dy$.

由积分定义知

$$\begin{aligned}G(x)&=\lim_{n\to\infty}\frac{1}{h}\sum_{k=1}^{n}F\Big(x+\frac{k}{n}h\Big)\frac{h}{n}\\&=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}F\Big(x+\frac{kh}{n}\Big).\end{aligned}$$

而 $F\big(x+\frac{kh}{n}\big)\equiv F_k(x)$ 是分布函数，从而，其凸线性组合 $G_n(x)\equiv\frac{1}{n}\sum\limits_{k=1}^{n}F\Big(x+\frac{kh}{n}\Big)$ 显然亦是分布函数，从而， $G(x)$ 是分布函数列 $G_n(x)$ 的极限函数.

如果能证明 $G(x)$ 是分布函数，则由特征函数列的极限定理知： $G_n(x)$ 的特征函数

$$\begin{aligned}g_n(t)&=\frac{1}{n}\sum_{k=1}^{n}e^{-i\frac{kht}{n}}\cdot f(t)\overset{n\to\infty}{\longrightarrow}f(t)\cdot\lim_{n\to\infty}\frac{1}{n}\cdot\frac{e^{-i\frac{ht}{n}}(1-e^{-iht})}{1-e^{-i\frac{ht}{n}}}\\&=f(t)\lim_{n\to\infty}\frac{1}{n}\cdot\frac{1-e^{-iht}}{e^{i\frac{ht}{n}}-1}=f(t)(1-e^{-iht})\lim_{n\to\infty}\frac{1}{n\Big(\frac{iht}{n}+\frac{(iht)^2}{2!n^2}+\cdots\Big)}\\&=\frac{1-e^{-iht}}{iht}\cdot f(t)\equiv g(t),\end{aligned}$$

其中 $g(t)$ 必是 $G(x)$ 的特征函数. 再由 $G(x)$ 的连续性、逆转公式以及积分的绝对收敛性得证

$$\begin{aligned}G(T+h)-G(T)&=\lim_{l\to\infty}\frac{1}{2\pi}\int_{-l}^{l}\frac{1-e^{-ith}}{it}e^{-itT}g(t)\,dt\\&=\frac{1}{2\pi h}\int_{-\infty}^{\infty}\Big(\frac{1-e^{-ith}}{it}\Big)^2e^{-itT}f(t)\,dt.\end{aligned}$$

最后我们证明 $G(x)$ 是分布函数，由 $F(x)$ 的非负性， $G(x)$ 显然是非负的；且由定义知 $G(x)$ 是连续函数 (积分限为变量的定积分是该变量的连续函数); 又由 $G'(x)=F(x+h)-F(x)\geqslant 0$ 及 G 的连续性知 $G(x)$ 单调不减；再由 $|F(x+t)|\leqslant 1$ 对任意 x 及 t 一致有界，从而可在积分号下取极限，有

$$\lim_{x\to+\infty}G(x)=\frac{1}{h}\int_0^h\lim_{x\to\infty}F(x+t)\,dt=1,\quad\lim_{x\to-\infty}G(x)=0.$$

故 $G(x)$ 的确是分布函数.

53. 提示：先证对一个区间上的均匀分布成立，再证均匀分布的凸线性组合成立，再取极限，即知对一般的分布密度也对. 本题就是富氏变换理论的 Riemann-Lebesgue 引理.

54. 实值特征函数可写为 $f(t)=\int_{-\infty}^{\infty}\cos tx\,dF(x)$, 于是

$$\begin{aligned}1-f(2t)&=2\int_{-\infty}^{\infty}\sin^2 xt\,dF(x)=2\int_{-\infty}^{\infty}(1-\cos xt)(1+\cos xt)dF(x)\\&\leqslant 4\int_{-\infty}^{\infty}(1-\cos xt)dF(x)=4-4f(t),\\1+f(2t)&=\int_{-\infty}^{\infty}(1+\cos 2xt)dF(x)=2\int_{-\infty}^{\infty}\cos^2 xt\,dF(x)\\&\geqslant 2\left(\int_{-\infty}^{\infty}\cos xt\,dF(x)\right)^2=2[f(t)]^2.\end{aligned}$$

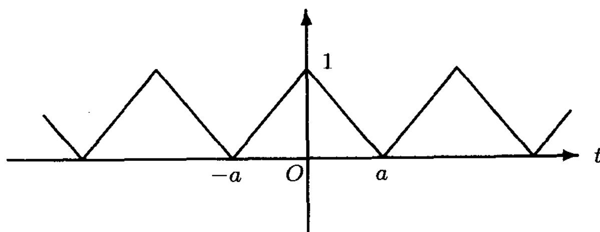
55. 提示：由 46 题， $\sum_{i=1}^{n}\xi_i$ 的特征函数为

$$\begin{aligned}\varphi(t)&=p^n(1-qe^{it})^{-n}=p^n\sum_m(-1)^m\frac{-n(-n-1)\cdots(-n-m+1)}{m!}(qe^{it})^m\\&=p^n\sum_m\frac{(m+n-1)!}{m!(n-1)!}q^m e^{itm}.\end{aligned}$$

由此可见 $P\left(\sum_{i=1}^{n}\xi_i=m\right)=\frac{(m+n-1)!}{m!(n-1)!}p^nq^m$.

56. 提示： $G(x)$ 是 $F(x)$ 与 $(-h,h)$ 上均匀分布的卷积；并利用 44 题的结果 (令 $a=h$).

57. 证. 由题中条件所规定的函数是以 $2a$ 为周期的对称函数. 将 $f(t)$ 展成富氏余弦级数



$$\begin{aligned}f(t)&=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos\frac{n\pi t}{a},\\a_0&=\frac{2}{a}\int_0^a\frac{a-x}{a}\,dx=1,\\a_n&=\frac{2}{a}\int_0^a\frac{a-x}{a}\cos\frac{n\pi x}{a}\,dx=-\frac{2}{n^2\pi^2}[(-1)^n-1]\\&=\begin{cases}0, & n=2k,\\ \dfrac{4}{\pi^2(2k-1)^2}, & n=2k-1,\ k=0,1,\cdots,\end{cases}\\f(t)&=\frac{1}{2}+\frac{4}{\pi^2}\left[\cos\frac{\pi t}{a}+\frac{1}{3^2}\cos\frac{3\pi t}{a}+\frac{1}{5^2}\cos\frac{5\pi t}{a}+\cdots\right]\\&=\frac{1}{2}+\frac{4}{2\pi^2}\left[e^{i\frac{\pi t}{a}}+e^{-i\frac{\pi t}{a}}+\frac{1}{3^2}e^{i\frac{3\pi t}{a}}+\frac{1}{3^2}e^{-i\frac{3\pi t}{a}}+\cdots\right]\\&=\sum_{n=-\infty}^{\infty}p_n e^{t\frac{n\pi}{a}i},\end{aligned}$$

其中 $p_0=\frac{1}{2}$, $p_{2k}=0$, $p_{-(2k-1)}=p_{2k-1}=\frac{2}{\pi^2}\frac{1}{(2k-1)^2}$, $k=1,2,\cdots$.

$$\sum_{n=-\infty}^{\infty} p_n=\frac{1}{2}+\frac{4}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2}=\frac{1}{2}+\frac{4}{\pi^2}\cdot\frac{\pi^2}{8}=1.$$

故 $f(t)$ 是对应于离散型随机变数

$$\begin{pmatrix} \frac{n\pi}{a} & ;\ n=0,\pm1,\cdots \\ p_n & \end{pmatrix}$$

的特征函数.

第 三 章

1. 证. 令 $B=\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n$. 又 $\overline{A}_n$ 表 A_n 的对立事件, 则由 §3.1 对偶原则得

$$1-P(B)=\lim_{k\to\infty}P\left(\bigcap_{n=k}^{\infty}\overline{A}_n\right)=\lim_{k\to\infty}\prod_{n=k}^{\infty}(1-P(A_n)).$$

由假设 $\sum_{n=1}^{\infty}P(A_n)=\infty$, 故无穷乘积 $\prod_{n=1}^{\infty}(1-P(A_n))=0$, 从而上式右方的极限等于 0, 亦即 $P(B)=1$.

2. 证. 必要性: 设 ξ_n 的分布是 F_n

$$\begin{aligned} E[f(|\xi_n|)]&=\int_{-\infty}^{\infty}f(|x|)]dF_n(x)=\int_{|x|>\varepsilon}f(|x|)dF_n+\int_{|x|\le\varepsilon}f(|x|)dF_n \\ &\leqslant \sup_{x\geqslant0}f(x)\int_{|x|>\varepsilon}dF_n+f(\varepsilon)\int_{|x|\le\varepsilon}dF_n \\ &\leqslant cP(|\xi_n|>\varepsilon)+f(\varepsilon). \end{aligned}$$

对任给的 $\eta>0$, 由于 $f(x)$ 的连续性, 存在 $\varepsilon>0$, 使 $f(\varepsilon)<\frac{\eta}{2}$.

由于 $\xi_n\to0$,(p), 存在 N, 使 $n>N$ 时

$$P(|\xi_n|>\varepsilon)<\frac{\eta}{2c}.$$

即 $n>N$ 时, $E[f(|\xi_n|)]<\eta$.

充分性:

$$\begin{aligned} E[f(|\xi_n|)]&=\int_{-\infty}^{\infty}f(|x|)]dF_n(x)=\int_{|x|>\varepsilon}f(|x|)dF_n+\int_{|x|\le\varepsilon}f(|x|)dF_n \\ &\geqslant f(\varepsilon)\int_{|x|>\varepsilon}dF_n=f(\varepsilon)p(|\xi_n|>\varepsilon). \end{aligned}$$

既然 $E[f(|\xi_n|)]\to0$, 而且对任何 $\varepsilon>0$, $f(\varepsilon)>0$, 故

$$P(|\xi_n|>\varepsilon)\to0,\ \text{即}\xi_n\to0,(\text{p}).$$

3. 提示: 利用特征函数的逆极限定理.

4. 证. $$e^{f(t)-1}=\frac{1}{e}\left(1+f(t)+\frac{1}{2!}f^2(t)+\cdots+\frac{1}{n!}f^n(t)+\cdots\right).$$
考虑 $$f_n(t)=\frac{1}{K_n}\left(1+f(t)+\frac{1}{2!}f^2(t)+\cdots+\frac{1}{n!}f^n(t)\right),$$
其中 $K_n=1+1+\frac{1}{2!}+\cdots+\frac{1}{n!}$. 由于 $f(t)$ 是特征函数，故 $f(t),f^2(t),\cdots,f^n(t)$ 也是，而 1 也是特征函数；既然 $f_n(t)$ 是它们的凸线性组合，所以也是特征函数.

另一方面 $e^{f(t)-1}$ 是连续的. 且

$$e^{f(t)-1}=\lim_{n\to\infty}f_n(t).$$

所以由逆极限定理 $e^{f(t)-1}$ 是特征函数.

5. 提示：如 ξ 有参数为 λ 的 Poisson 分布，则 $E\xi=D\xi=\lambda;\xi_n$ 可看成 n 个独立的具有参数为 1 的 Poisson 分布的随机变数的和.

6. 提示：由第一章题 14(iii),$q_k=C_{n-k}^{N+n-k-2}/C_n^{N+n-1}$, 故

$$\begin{aligned}q_k&=(N-1)n(n-1)\cdots(n-k+1)/(N+n-1)\cdots(N+n-k-2+1)\\&=\frac{N-1}{N+n-1}\cdot\frac{n}{N+n-2}\cdot\cdots\cdot\frac{n-k+1}{N+n-k-1}.\end{aligned}$$

注意条件 $N\to\infty,n\to\infty,\frac{n}{N}\to\lambda$, 即得 $q_k\to\lambda^k/(1+\lambda)^{k+1}$.

7. 提示：利用车贝晓夫不等式.

8. 提示：利用第二题的结果，取 $f(x)=\frac{x^2}{1+x^2}$ 即可.

9. 提示：注意 ξ_n 有 $N(\theta^n,\sqrt[4]{n}/\sqrt{2})$ 分布，而 $\sum\limits_{n=1}^{\infty}\frac{1}{n^2}\frac{\sqrt{n}}{2}<\infty$.

12. 柯尔莫哥洛夫条件不满足，但马尔科夫条件满足.

13. 提示：利用不等式，对 $m<n$ 有

$$\frac{1}{n^2}\sum_{k=1}^{n}DX_k\leq\frac{1}{n^2}\sum_{k=1}^{m}DX_k+\sum_{k=m+1}^{n}DX_k/k^2.$$

14. 令

$$\eta_{n,k}=\begin{cases}\xi_k, & |\xi_k|\leqslant n\delta\\ 0, & |\xi_k|>n\delta\end{cases}\quad(\delta>0)\quad k=1,2,\cdots,n.$$

由所给条件知

$$b_k=\int_{-\infty}^{\infty}|x|\,dF_k(x)\quad k=1,2,3,\cdots$$

存在而且 b_k 有界，设其界为 M. 由车贝晓夫不等式：

$$P\left[\left|\frac{1}{n}\sum_{k=1}^{n}\eta_{n,k}-\frac{1}{n}\sum_{k=1}^{n}E\eta_{n,k}\right|\geqslant\varepsilon\right]\leqslant\frac{M\delta}{\varepsilon^2}.\tag{1}$$

另一方面

$$\sum_{k=1}^{n}P(\xi_k\neq\eta_{n,k})=\sum_{k=1}^{n}P(|\xi_k|>n\delta)\leqslant\frac{1}{\delta}\max_{1\leqslant k<\infty}\int_{|x|\geqslant n\delta}|x|dF_k(x).\tag{2}$$

综合 (1),(2) 式易推知

$$\lim_{n\to\infty}P\left(\frac{1}{n}\left|\sum_{k=1}^{n}(\xi_k-E\xi_k)\right|\geqslant2\varepsilon\right)=0,$$

亦即 ξ_k 服从大数定律.

16. 能直接验证麟德伯尔格条件成立.

17. (i) 成立. (ii) 成立.

18. 提示：利用 $[0,1]$ 上的独立均匀分布序列服从强大数定律.

20. 证. 考虑 $\{\xi_n\}, \xi_n$ 的分布是 $\chi^2(1)$. 则 $E\xi_n = 1$, $E\xi_n^2 = 3$, $D\xi_n = 2$. 所以 ξ_n 服从中心极限定理. 又 $\sum\limits_{k=1}^{n} \xi_k$ 的分布是 $\chi^2(n)$.

$$
\begin{aligned}
& P\left(\frac{1}{\sqrt{2n}}\sum_{k=1}^{n}(\xi_k - 1) \leqslant t\right) = P\left(\sum_{k=1}^{n}\xi_k < n + \sqrt{2n}t\right) \\
&= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^{n+t\sqrt{2n}} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}\, dz \\
&= \frac{1}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\left(\frac{n}{2}\right)^n} \int_0^{1+t\sqrt{2/n}} z^{\frac{n}{2}-1} e^{-\frac{nz}{2}}\, dz \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{z^2}{2}}\, dz.
\end{aligned}
$$

21. 提示：若用 ξ_i 表示第 i 次试验的结果，即 $P(\xi_i = 1) = p_i$, $P(\xi_i = 0) = q_i$, 则 $s_n = \sum\limits_{i=1}^{n} \xi_i$. 只须验证 $\{\xi_i\}$ 是服从中心极限定理的即可.

22. (i) 当废品率为 0.005 时，检查 200 件产品，出现 4 件废品是一小概率事件 (概率约为 0.015), 故不可信.

 (ii) 13.

 (iii) 110.

23. 1598700.

第 四 章

1. 三个状态，是遍历的， $p_1 = \frac{2}{5}, p_2 = \frac{13}{35}, p_3 = \frac{8}{35}$.

2. $E = (1, 2, \cdots, 6)$; $p_{jk} = \frac{1}{6}, (k > j)$; $p_{jk} = \frac{k}{6}, (k = j)$; $p_{jk} = 0, (k < j)$; $k, j = 1, 2, \cdots, 6$.

3.
$$
\begin{aligned}
& P(\xi_{t_1} = j_1, \cdots, \xi_{t_k} = j_k;\ 0 < t_1 < \cdots < t_k) \\
&= \begin{cases} \dfrac{(\lambda t_1)^{j_1} e^{-\lambda t_1}}{j_1!} \displaystyle\prod_{u=1}^{k-1} \dfrac{[\lambda(t_{u+1} - t_u)]^{j_{u+1} - j_u}}{(j_{u+1} - j_u)!} e^{-\lambda(t_{u+1} - t_u)}, & \text{如} 0 \leqslant j_1 \leqslant j_2 \leq \cdots \leqslant j_k \\ 0, & \text{否则} \end{cases}
\end{aligned}
$$

4. $(\xi_{t_1}, \cdots, \xi_{t_k})$ $(0 < t_1 < \cdots < t_k)$ 的联合密度为

$$
\frac{1}{\sqrt{2\pi t_1}} \exp\left(-\frac{x_1^2}{2t_1}\right) \prod_{u=1}^{k-1} \frac{1}{\sqrt{2\pi(t_{u+1} - t_u)}} \exp\left(-\frac{(x_{u+1} - x_u)^2}{2(t_{u+1} - t_u)}\right)
$$

5. $B_\eta(t) = \sum\limits_{i,j} a_i \bar{a}_j B_\xi(\tau + s_i - s_j)$; $F_\eta(x) = \int_{-\infty}^{x} \left|\sum\limits_{k} a_k e^{i\lambda s_k}\right|^2 dF_\xi(\lambda)$.

第 五 章

1. $\frac{1}{2}\left(\max\limits_{1 \leqslant i \leqslant n} x_i + \min\limits_{1 \leqslant i \leqslant n} x_i\right)$ 较有效.

2. $\frac{4}{3}\max\limits_{1\leqslant i\leqslant 3} x_i$ 较有效.

3. $k_1=\frac{1}{3}$, $k_2=\frac{2}{3}$.

5. 凡是满足不等式

$$\max_{1\leqslant i\leqslant n} x_i-\frac{1}{2}\leqslant \widehat{\theta}(x_1,\cdots,x_n)\leq \min_{1\leqslant i\leqslant n} x_i+\frac{1}{2}$$

的估计值 $\widehat{\theta}$ 都是最大似然估值，例如，$\frac{1}{2}\Big(\max\limits_{1\leqslant i\leqslant n} x_i+\min\limits_{1\leqslant i\leqslant n} x_i\Big)$ 就是一个.

提示：直接按照最大似然估值的定义求.

6. $\max\limits_{1\leqslant i\leqslant n} x_i$(直接按定义求).

7. (i)$\overline{x}$. (iii) $\overline{x}$.

(iv) 将 $x_1,\cdots,x_n$ 按大小排成 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$.

当 $n=2k+1$ 时. $\widehat{\theta}=x_{(k+1)}$ 是最大似然估值；

当 $n=2k$ 时，一切满足 $x_{(k)}<\widehat{\theta}\leqslant x_{(k+1)}$ 的 $\widehat{\theta}$ 都是最大似然估值. 当 $x_{(k)}=x_{(k+1)}$ 时，令 $\widehat{\theta}=x_{(k)}$.

8. $\widehat{\theta}_1=\min\limits_{1\leqslant i\leqslant n} x_i$, $\widehat{\theta}_2=\frac{1}{n}\sum\limits_{i=1}^{n}\Big(x_i-\min\limits_{1\leqslant i\leqslant n} x_i\Big)$.

9. 提示：要检查两个正态分布的期望值是否相等，利用 §5.5 (二)3° 所给的检验法.

$$T=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\cdot\frac{\overline{\xi}-\overline{\eta}}{\sqrt{n_1s_1^2+n_2s_2^2}}=0.312.$$

自由度 $n_1+n_2-2=21$. 查自由度为 21 的 t- 分布表，信度取 1% 得 $2.83>0.312$, 所以按 1% 信度，两店的豆可以看作是同一种类的.

10. 提示：用 §5.5(二) 1° 的检查法.

$$t=\sqrt{n-1}\frac{\xi-a_0}{s}=2.10.$$

取信度 5%, 查自由度为 23 的 t- 分布表，得 $2.07<2.10$, 故不能认为期望值是 12100.

11. 提示：解法同第 9 题，令

$$T=4.42.$$

查自由度为 18 的 t- 分布表，信度取 1%, 得 $2.878<4.42$. 故二品种产量有显著差别.

12. 提示：用 §5.5(二) 4° 的检查法.

$$F=1.03.$$

取 5% 的信度，查自由度为 5; 8 的 F- 分布表，得 $3.69>1.03$. 故可认为二种口径方差无差异.

13. 提示：用 §5.5(三) 中的 χ^2- 检验法.

首先算得 $\overline{x}=0.805$, 现要检查过路汽车的辆数是否是 $\lambda=0.805$ 的普阿松分布.

数 值 表 一

供正态分布查用.

函数　　$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$ 数值表 [1]

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.0000						
0.01	0.0040	0.31	0.1217	0.61	0.2291	0.91	0.3186
0.02	0.0080	0.32	0.1255	0.62	0.2324	0.92	0.3212
0.03	0.0120	0.33	0.1293	0.63	0.2357	0.93	0.3238
0.04	0.0160	0.34	0.1331	0.64	0.2389	0.94	0.3264
0.05	0.0199	0.35	0.1368	0.65	0.2422	0.95	0.3289
0.06	0.0239	0.36	0.1406	0.66	0.2454	0.96	0.3315
0.07	0.0279	0.37	0.1443	0.67	0.2486	0.97	0.3340
0.08	0.0319	0.38	0.1480	0.68	0.2517	0.98	0.3365
0.09	0.0359	0.39	0.1517	0.69	0.2549	0.99	0.3389
0.10	0.0398	0.40	0.1554	0.70	0.2580	1.00	0.3413
0.11	0.0438	0.41	0.1591	0.71	0.2611	1.01	0.3438
0.12	0.0478	0.42	0.1628	0.72	0.2642	1.02	0.3461
0.13	0.0517	0.43	0.1664	0.73	0.2673	1.03	0.3485
0.14	0.0557	0.44	0.1700	0.74	0.2703	1.04	0.3508
0.15	0.0596	0.45	0.1736	0.75	0.2734	1.05	0.3531
0.16	0.0636	0.46	0.1772	0.76	0.2764	1.06	0.3554
0.17	0.0675	0.47	0.1808	0.77	0.2794	1.07	0.3577
0.18	0.0714	0.48	0.1844	0.78	0.2823	1.08	0.3599
0.19	0.0753	0.49	0.1879	0.79	0.2852	1.09	0.3621
0.20	0.0793	0.50	0.1915	0.80	0.2881	1.10	0.3643
0.21	0.0832	0.51	0.1950	0.81	0.2910	1.11	0.3665
0.22	0.0871	0.52	0.1985	0.82	0.2939	1.12	0.3686
0.23	0.0910	0.53	0.2019	0.83	0.2967	1.13	0.3708
0.24	0.0948	0.54	0.2054	0.84	0.2995	1.14	0.3729
0.25	0.0987	0.55	0.2088	0.85	0.3023	1.15	0.3749
0.26	0.1026	0.56	0.2123	0.86	0.3051	1.16	0.3770
0.27	0.1064	0.57	0.2157	0.87	0.3078	1.17	0.3790
0.28	0.1103	0.58	0.2190	0.88	0.3106	1.18	0.3810
0.29	0.1141	0.59	0.2224	0.89	0.3133	1.19	0.3830
0.30	0.1179	0.60	0.2257	0.90	0.3159	1.20	0.3849

[1] 更多的数值表见参考书 [33], [47].

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1.21	0.3869	1.56	0.4406	1.91	0.4719	2.52	0.4941
1.22	0.3888	1.57	0.4418	1.92	0.4726	2.54	0.4945
1.23	0.3907	1.58	0.4429	1.93	0.4732	2.56	0.4948
1.24	0.3925	1.59	0.4441	1.94	0.4738	2.58	0.4951
1.25	0.3944	1.60	0.4452	1.95	0.4744	2.60	0.4953
1.26	0.3962	1.61	0.4463	1.96	0.4750	2.62	0.4956
1.27	0.3980	1.62	0.4474	1.97	0.4756	2.64	0.4959
1.28	0.3997	1.63	0.4484	1.98	0.4761	2.66	0.4961
1.29	0.4015	1.64	0.4495	1.99	0.4767	2.68	0.4963
1.30	0.4032	1.65	0.4505	2.00	0.4772	2.70	0.4965
1.31	0.4049	1.66	0.4515	2.02	0.4783	2.72	0.4967
1.32	0.4066	1.67	0.4525	2.04	0.4793	2.74	0.4969
1.33	0.4082	1.68	0.4535	2.06	0.4803	2.76	0.4971
1.34	0.4099	1.69	0.4545	2.08	0.4812	2.78	0.4973
1.35	0.4115	1.70	0.4554	2.10	0.4821	2.80	0.4974
1.36	0.4131	1.71	0.4564	2.12	0.4830	2.82	0.4976
1.37	0.4147	1.72	0.4573	2.14	0.4838	2.84	0.4977
1.38	0.4162	1.73	0.4582	2.16	0.4846	2.86	0.4979
1.39	0.4177	1.74	0.4591	2.18	0.4854	2.88	0.4980
1.40	0.4192	1.75	0.4599	2.20	0.4861	2.90	0.4981
1.41	0.4207	1.76	0.4608	2.22	0.4868	2.92	0.4982
1.42	0.4222	1.77	0.4616	2.24	0.4875	2.94	0.4984
1.43	0.4236	1.78	0.4625	2.26	0.4881	2.96	0.4985
1.44	0.4251	1.79	0.4633	2.28	0.4887	2.98	0.4986
1.45	0.4265	1.80	0.4641	2.30	0.4893	3.00	0.49865
1.46	0.4279	1.81	0.4649	2.32	0.4898	3.20	0.49931
1.47	0.4292	1.82	0.4656	2.34	0.4904	3.40	0.49966
1.48	0.4306	1.83	0.4664	2.36	0.4909	3.60	0.499841
1.49	0.4319	1.84	0.4671	2.38	0.4913	3.80	0.499928
1.50	0.4332	1.85	0.4678	2.40	0.4918	4.00	0.499968
1.51	0.4345	1.86	0.4686	2.42	0.4922	4.50	0.499997
1.52	0.4357	1.87	0.4693	2.44	0.4927	5.00	0.49999997
1.53	0.4370	1.88	0.4699	2.46	0.4931		
1.54	0.4382	1.89	0.4706	2.48	0.4934		
1.55	0.4394	1.90	0.4713	2.50	0.4938		

数值表二：χ^2 分布表

设 χ^2 的密度由 §2.8(30) 给出，对已给的 p, $0 < p < 100$, 满足下式的常数 χ_p^2 由下表可以查到，因而 χ^2 大于 χ_p^2 的概率为 $p\%$.

$$P(\chi^2 > \chi_p^2) = \frac{p}{100}.$$

自由度	χ_p^2 作为 n 与 p 的函数													
n	$p=99$	98	95	90	80	70	50	30	20	10	5	2	1	0.1
1	0.000	0.001	0.004	0.016	0.064	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	0.020	0.040	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

数 值 表 三：t 分布表

设 t 的密度由 §2.8(33) 给出，对已给的 p, $0 < p < 100$, 满足下式的常数 t_p 由下表可以查到，因而 t 的绝对值大于 t_p 的概率为 $p\%$.

$$P(|t| > t_p) = \frac{p}{100}.$$

自由度	t_p 作为 n 与 p 的函数												
n	$p = 90$	80	70	60	50	40	30	20	10	5	2	1	0.1
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.073	6.314	12.706	31.821	63.657	636.619
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.589
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.126	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

论 随 机 性

1. 随机性与必然性的相互交替

在某些条件下，一定出现 (或一定不出现) 的事件，称为必然事件；或者说，这事件具有必然性. 在某些条件下，可能出现，也可能不出现的事件，称为随机事件 (也称为偶然事件); 或者说，这事件具有随机性 (偶然性). 必然与偶然，都是相对于条件而言的. 在 101.325kPa 气压及 100° 温度时，“水必沸腾” 是必然事件；“明年某河流将泛滥成灾” 是随机事件.

事物的发展，是多层次的随机事件与必然事件相互交替和相互作用的过程. 有些人说，发展过程中必然性 (或决定性) 是主导的，随机性是次要的；另一些人则反之. 二者皆失之偏颇，不符合客观实际.

种子随机地扎根于某地 (1 层), 默默地度过一段时间后，必然破土而出 (2 层), 这时它只有一根主干. 到了某一时刻，它分成几支，分支的时刻和支数都是随机的 (3 层). 每一支又平静地生长 (4 层), 再分支 (5 层), 如此继续. 若干年后，它或遭意外，或者枯萎，这死亡的时刻和原因又是随机的.

婴儿呱呱堕地，他的出生充满了偶然性 (1 层). 他平静地度过童年 (2 层).18 岁时，他站在人生道路的分支点上，在就业和升学间进行随机选择 (3 层). 过了这一关，他可以安心地工作和学习相当长时间 (4 层), 然后又一次站在分支点上 (5 层) …… 直至某一偶然的时刻，来到偶然的地点，出于偶然的原因，结束他那偶然的生命.

邱吉尔说：“一个人活得愈长，他就愈认识到一切取决于机会. 任何人哪怕只要回顾一下 10 年前的经历，他就会看到某些本身毫不重要的细小事件，实际上都左右了他的全部命运和前程. ” 他所说的细小事件，大都集中在分支点上. “却顾所来径，苍苍横翠微”. 只有回过头来，才能认识它们的巨大作用，于是大吃一惊而感叹万千.

天文学家还不能确切告诉我们地球是怎样诞生的. 行星 (甚至太阳) 的形成有很大的偶然性. 地球也必定经历了许多分支点. 地质学家说地球至少发生过 6 次生物大规模灭绝 [1], 最著名的一次是 0.6 亿年前的恐龙灭绝. 如果没有那次灾变，今天的地球也许还是恐龙世界. 生物灭绝的原因有种种假说：超新星爆发、行星撞击、太阳耀斑爆发、海平面变化、温度变化等等，无一不充满偶然性. 多么危险的地球！地球如此，极而言之，整个宇宙也是偶然的. 导致混沌初开的大爆炸，星系的形成与分布，生命的出现，行星上有多少种动物和植物，每一种中有多少个体，都含有许许多多偶然的因素.

每个国家的历史也有许多分支点，如果在分支点上换另一种选择，历史便当另写. 设若在鸿门宴上项羽杀了刘邦，便很可能没有汉朝，没有刘皇叔 (刘备), 没有大家津津乐道的诸葛亮.

说了这么多的偶然性，难道 “人是要死的”,“明天太阳还会升起” 也有偶然性吗？否！这是千真万确的必然事件. 不过，生命之于身体，正如飞之于飞机，飞是飞机的功能，生命也是身体的功能. 生命是暂存的，而构成身体的物质是永恒的；生命只是这些物质在 2 个分支点 (出生与死亡) 之间的一种运动形式，而在这 2 点之间，运动是相对平稳的、缓变的，必然性居主导地位，它必然地要走向下一分支点，即死亡. 但这分支点何时到来、如

何到来以及以后这些物质如何继续运动，则是随机的. 同样，太阳东起西落，也只是地球在 2 个分支点 (从稳定运行到失稳) 间的决定性运动，在失稳以后，还能说太阳明天升起吗？

关于太阳升起问题，拉普拉斯做过研究：假设太阳已接连升起 n 次，问它还会升起 1 次的概率是多少？他的答案是. $(n+1)/(n+2)$. 当 n 很大时这几乎等于 1, 所以我们对明天的安全可以放心. 不过这答案只适用于地球处于上述 2 个分支点之间时；不然，若 $n\to\infty$ 这概率 $\to 1$, 而这是荒谬的，因为太阳系决不会永久稳定下去，总有一天它会瓦解.

上述众多事实说明了事物发展的一般法则：发展过程是偶然与必然的相互交替和相互作用. 在发展道路上有许多分支点，在相邻 2 点之间，发展是相对稳定的、量变的、合乎逻辑的，必然性起着主导作用；但这里也可能有次要的分支点和次要的随机因素. 分支点的到来是随机的，在分支点上，发展前途面临多种选择，必须选择其一而尽弃其余，这时运动是不稳定的、突变的，随机性起主导作用. 一旦选定以后，发展又趋于稳定，直到下一分支点，如此继续，如图 1 所示. 事物必需从 $A_0A_1B_1C_1,\cdots,A_0A_1B_3C_7$ 7 条道路中选择 1 条，尽管这 1 条未必比其他 6 条都优越. 由此可见，我们的宇宙史、国家史、家庭史、个人史，都只是许许多多可供选择中的一种，它们全都富含随机性；如果说得偏激点，它们全是随机的.

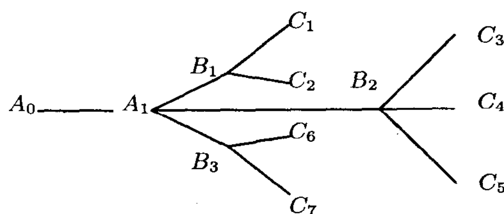


图 1

偶然性来自何方？来自事物内部的变化，来自外环境的碰撞. 构成事物的各种物质、各种因素以及支配它们的各种力经常处于运动之中. 如果各自的运动处于允许范围之内，那么事物整体的发展是稳定的、缓变的；一旦某些主要部分的运动超越甚至远离允许范围时，整个事物就会发生突变，而那些部分超越则是随机的. 可以设想一个混乱社会最后如何随机转变的情形以增加我们的想象力. 事物处于外环境中，它经常与其他事物发生关系，称之为碰撞. 其作用小则可影响事物的量 (如速度、大小), 大则可影响事物的质 (如方向、性质). 人类社会是一个极富于随机性的碰撞世界：亲属、朋友、师生、同事的结合是随机的，人与人，人与事，人与物的相互作用，形势的影响，都可视为碰撞. 即使是站着不动的植物，也要受到自然、环境及动物的种种碰撞. 不过有些碰撞 (如上街遇到的许多行人) 未起作用，为人们所忽视，而起作用的 (如车祸), 则可使人终生难忘. 外部的碰撞可以促进内因的变化，内外是互为影响的，不能截然分开.

结构单纯，质量和体积巨大的物体有比较稳定的内环境，如果它又远离其他大物体，从而外部碰撞甚少，那么，相对于这种物体，偶然性便很小，分支点之间的距离便很长，这就是为什么天体运行基本上遵循必然法则的原因. 如果我们寿命长达百亿年，我们就会看到天体经过许多分支点而作随机运动，那时就不会说天体运动是必然的了. 另一方面，极小的粒子在碰撞下极不稳定，微观世界所以基本上是偶然的世界，这是原因之一.

以上讨论了偶然与必然的交替，提出了分支点的概念，还谈了偶然性的来源. 下面考虑偶然性与必然性的关系.

随机性大都出现在分支点附近，但这并不意味着二相邻分支点间没有随机性，一些次要的偶然事件，次要的分支点经常在这里出现；此外，我们在前面还谈到正是在这段时间内孕育着新的随机性. 在许多场合，偶然受到必然的限制. 必然性给它划定了一个范围，

随机选择只能在此范围内进行. 例如，婴儿的性别虽是偶然的，但只有 2 种可能.

随着条件的改变，必然事件与偶然事件可互相转化. 向一巨大的目标射击，击中是必然的；现在设想目标不断缩小，起初击中还是必然的，但小到某一限度后，击中目标便成为偶然事件了. 这提示人们可以设想有一临界值或临界区存在. 这是必然转化为偶然的例. 另方面，在正常交通情况下，车祸是偶然的，但在禁止车辆通行的路段，车祸不可能发生，偶然化为必然.

彭加来说："最大的机遇莫过于一个伟人的诞生"[2]. 其所以如此，一是由于某人的诞生是一系列随机事件的复合：父母，祖父母，外祖父母的结合，异性的 2 个生殖细胞的相遇，而这 2 个细胞又必须含有某些产生天才的因素. 另一是婴儿出生以后，各种偶然遭遇在整体上必须有利于他的成功，他所处的时代，他所接受的教育、他的各项活动、他所接触的人与事与物，都需为他提供好的机会. 所以，某个特定的人要成为伟人，可能性是极小的.

虽然如此，各时代仍然伟人辈出. 一个人成功的概率虽极小，但几十亿人中总有佼佼者，这就是所谓"必然寓于偶然之中"的一种含义. 数学中称之为"小概率原理". 设某试验中出现事件 A 的概率为 ε, 不管 $\varepsilon > 0$ 如何小，如果把这试验不断独立地重复下去，那么 A 迟早必然会出现 1 次，从而也必然会出现任意多次. 这是因为，第 1 次试验中 A 不出现的概率为 $1-\varepsilon$, 前 n 次 A 都不出现的概率为 $(1-\varepsilon)^n$, 因之前 n 次试验中 A 至少出现 1 次的概率为 $1-(1-\varepsilon)^n$. . 当 $n \to \infty$ 时这概率趋于 1, 这表示 A 迟早出现 1 次的概率为 1. 出现 A 以后，把下次试验当作第 1 次，重复上述推理，可见 A 必然再次出现，如此继续，可知 A 必然出现任意多次. 应用此原理于伟人问题，一个人成为伟人的概率 ε 固然非常小，但千百万人中至少有一伟人就几乎是必然的了.

"必然寓于偶然之中"的另一含义是大数定律，它的特殊情形是频率的稳定性，即频率趋于概率. 设某试验中事件 A 出现的概率为 $p > 0$, 将此试验独立地重复 n 次，其中 A 出现了 m 次，于是频率为 $\frac{m}{n}$. 根据大数定律，当 $n \to \infty$ 时，必然有 $\frac{m}{n} \to p$. 因此，当 n 充分大时，得. $m \approx np$.

我们不能确切预知一个婴儿的性别，只知他是男性的概率为 1/2; 但由上述定律，我们可以断言，100 万婴儿中，约有 1/2 即 50 万个男婴，这几乎是必然的.

2. 随机试验

每个随机事件都联系于一随机试验；后者完整地刻画了一族随机事件.

固定条件组 $C = (c_1, c_2, \cdots, c_n)$, 其中 c_i 代表第 i 个条件. 如前所述，在 C 下，一定出现的事件 A 称为 (相对于 C 的) 必然事件. 其模型为

$$\left.\begin{matrix} c_1 \\ \vdots \\ c_n \end{matrix}\right\} \Longrightarrow A. \tag{1}$$

随机试验 (简称试验) E 的模型则为

$$\left.\begin{array}{l} c_1 \\ \vdots \\ c_n \end{array}\right\} \Longrightarrow \begin{cases} (A_1, p_1) \\ \vdots \\ (A_m, p_m). \end{cases} \tag{2}$$

它表示: 在条件组 $C=(c_1,c_2,\cdots,c_n)$ 下, 可能出现的结果是 $A_1,\cdots,A_m$ 中之一, 一次试验中出现 A_i 的概率为 $p_i, 1>p_i>0, p_1+p_2+\cdots+p_m=1.$, 有时也称 A_i 为状态.

由于 A_i 在一次试验中可能出现 (概率为 p_i), 也可能不出现 (概率为 $1-p_i$), 所以 A_i 是一随机事件, $(i=1,\cdots,m)$. 我们假定 $m>1$ 以保证 E 的随机性. 如果允许 $m=1$, 这时 (2) 便化为 (1), 于是必然事件可看成为概率是 1 的随机事件.

例 1 无偏倚地 (条件 c_1) 扔一枚正常的 (c_2) 硬币, 登记朝上的面. 2 个可能的状态及概率分别为

$$(A_1,p_1)=(\text{正},1/2);\ (A_2,p_2)=(\text{反},1/2).$$

登记婴儿性别也同此试验, 只需理解 " 正, 反 " 为 " 女, 男 ".

例 2 军事家克劳塞维茨说: "人类的任何活动都不像战争那样给偶然性这个不速之客留有这样广阔的活动天地, 因为没有一种活动像战争这样从各方面和偶然性经常接触." 天气, 一次射击的效果, 一个士兵的暴露, 司令官选择策略, ⋯⋯ 都有偶然性, 我们可能把甲乙双方进行的战争看成一随机试验, 条件 C 由双方目前的军事, 经济等情况组成, 可能的结果有 4:

$$A_1\text{— 甲胜};\quad A_2\text{— 乙胜};\quad A_3\text{— 和局};\quad A_4\text{— 两败俱伤}.$$

在战争初期, 很难确定 A_i 的概率, 它随事态发展而逐步明确. 以后会看到, 这是归一型的随机试验.

条件 C 决定各状态的概率 $p_j, j=1,\cdots,m,\ p_j$, 可看成 C 的泛函. 如何从 C 求出 p_j 是一非常重要而一般又很困难的问题. 多数情况 p_j 是 C 的非线性泛函, 除了 "各结果等可能" 或 "几何概率" 等几种特殊情况外, 并无求 p_j 的一般方法, 需要具体问题具体解决. 但对可重复试验, 如果精确度的要求不太高, 如前面谈过的可以用频率来近似地求出 p_j. 由于频率是客观的存在, 所以对可重复试验, 事件的概率是客观的, 不能由人主观地任意赋予 [3].

我们说随机试验 E 是可重复的, 是指在相同的条件 C 下, 可将 E 不断地独立地进行下去. 上述例 1 中试验是可重复的, 人们可将此硬币不断扔掷. 所谓重复可如下理解: 或将 E 在 C 下独立进行 n 次; 或取 n 个与 E 相同的试验 $E_1,E_2,\cdots,E_n$ 在 C 下同时各进行 1 次, 或取 $\ell(\leqslant n)$ 个与 E 相同的试验在 C 下各做若干次, 使总次数为 n.

引进时间 t, 便得到随机试验的动态模型:

$$\left.\begin{array}{l} c_1(t) \\ \vdots \\ c_n(t) \end{array}\right\} \Longrightarrow \begin{cases} (A_1, p_1(t)) \\ \vdots \\ (A_m, p_m(t)). \end{cases} \tag{3}$$

其中 $1 > p_i(t) \geqslant 0, p_1(t) + \cdots + p_m(t) = 1$. 又 $c_i(t), p_i(t)$ 分别表示在时刻 t 的第 i 个条件及 A_i 的概率. 在 t 时，也许减少了、也许新增了一些条件，所以 n 也应依赖于 t, 即 $n = n(t)$; 同理， $m = m(t)$. 不过为了记号简单，把 t 略去了. 现在还允许 $p_i(t) = 0$, 以便表示在 t 时 A_i 不可能出现，尽管在别的时刻 A_i 可能出现.

以 e 表示试验结束的时刻， $0 < e \leq \infty$. 全体随机试验可分成互不相交的 3 种类型.

a. 归一型：存在一个状态 A_k, 使

$$\lim_{t \to e} p_k(t) = 1. \tag{4}$$

这表示：当试验接近尾声时， A_k 出现的概率越来越大，最后成为必然事件，于是偶然性逐渐转化为必然性. 存在一临界时刻 a, 使

$$p_k(t) \geqslant 90\%, \text{ 一切} t \in [a, e).$$

这表示自 a 而后，出现 A_k 的概率已不小于 0.9. 称 $[a, e)$ 为转化区间，其长 $e - a$ 随问题而定. 上述例 2 便是归一型的. 例如，到 1944 年冬，二次大战中日本失败已成为定局. 许多比赛也是归一型随机试验.

b. 分散型：存在极限

$$\lim_{t \to e} p_i(t) \equiv q_i \geqslant 0, \text{ 一切} i; \tag{5}$$

而且至少有 2 个状态 A_k, A_j, 使

$$q_k > 0, \ q_j > 0. \tag{6}$$

这表示，即使试验临近结束，仍然不能断定哪一状态必然出现，从而随机性贯彻始终. 例 1 属于分散型，因为 $p_1(t) \equiv p_2(t) \equiv 1/2$.

又如：在群体遗传学中， Hardy-Weinberg 定律表示：对 3 种基因型 AA, Aa, aa, 第 t 子代取这些型的概率分别为

$$p_t(AA) = p^2, \ p_t(Aa) = 2pq > 0, \ p_t(aa) = q^2.$$

它们与 $t = 1, 2, 3, \cdots$ 无关， p, q 为正常数. 故基因型试验是分散型的.

一般地，如在时刻 $t = 1, 2, 3, \cdots$ 将同一试验重做 1 次，假定这些试验彼此无关地独立进行，则 $p_i(t) \equiv q_i$ 与 t 无关， $i = 1, 2, \cdots, m$. 因试验是随机的，至少有 2 个非 0 的 q_i, 于是这一列试验是分散型的. 由此可见，分散型试验非常多.

c. 振动型：至少有一状态 k, 使极限 $\lim_{t \to e} p_k(t)$ 不存在；也就是说，

$$\overline{\lim_{t \to e}} p_k(t) > \underline{\lim_{t \to e}} p_k(t). \tag{7}$$

作为一例，考虑飞机飞行， A_1 表无事故， A_2 表有事故， $p_1(t)$ 是第 t 日飞行无事故的概率. 由于天气影响飞行，好日子事故少. 以 s_i 表好日子， t_j 表坏日子，有

$$\overline{\lim_{t \to b}} p_1(t) = \lim_{s_i \to b} p_1(s_i) > \lim_{t_j \to b} p_1(t_j) = \underline{\lim_{t \to b}} p_1(t),$$

因而飞行可看成振动型随机试验.

上述 3 种类型已穷尽一切随机试验，每一试验必定也只能属于三者之一. 下面将看到，这种分类对讨论拉普拉斯的决定论很有帮助.

概率分布 $P(t)=(p_1(t),\cdots,p_m(t))$ 的熵定义为

$$I(P(t))=-\sum_{i=1}^{m}p_i(t)\log p_i(t).$$

它是此分布的偶然性大小 (或称为混乱程度) 的量度，也可把它看成定义在分布所成集合上的泛函. $I(P(t))$ 越大，$P(t)$ 的偶然性也越大. 可以证明，当 $P(t)$ 为均匀分布 $(1/m,\cdots,1/m)$ 时，熵取极大值 $\log m$. 容易看出：

对归一型试验， $\lim\limits_{t\to e}I(P(t))=0$;

对分散型试验， $\lim\limits_{t\to e}I(P(t))>0$;

对振动型试验， $\lim\limits_{t\to e}I(P(t))$ 不存在.

故用熵的极限，也可区分这 3 类试验.

更广泛的随机事件，是一系列随机试验串联的结果，例如上述伟人的诞生；另一些则是由许多试验并联而产生，如碰撞或巧遇. 最一般地，可以同时考虑串联与并联，于是得到复合随机试验的图式 (图 2): 图中每一 E 代表一随机试验.

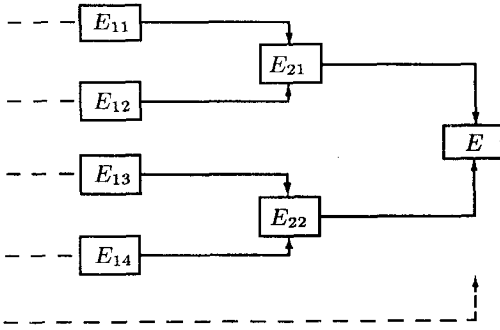


图 2

巧遇是一种复合试验，它使世界丰富多采，令人惊奇不已. 巧遇是小概率事件，它一般不会出现；而一旦出现，就可能创造奇迹. 例如：美国有 15 人将参加晚 7 时 15 分的排练，他们都因不同原因而迟到；那晚排练室被人放了定时炸弹， 7 时 25 分爆炸，这 15 人全因迟到而免于难，真是幸运之至. 又如：地球上生命的出现也是一系列巧合的结果，它要求地球离太阳不远不近，质量不大不小，恰好都是现在那样，还要求有一个大得出奇的卫星 (月亮), 等等. 由于这种巧合的概率太小，不少人认为人类在宇宙中是孤独的，不存在地球外的生命. 人的一生有许多巧遇，其中部分是好机会，另一些则是倒霉事. 聪明人善于抓住、利用甚至制造好机会，同时尽量避免坏机会. 衡量 “巧” 的尺度是概率的 “小”. 统计学家利用小概率事件来作假设检验：在假设 H 下设计一个小概率 (譬如 1%) 事件 A, 在一次试验中，这事件一般不会出现；但如果它居然出现了，便使人不得不怀疑假设 H 的正确性，因之否定 H. 在一些复合试验中，偶然性的后果可被层层放大. 例如，某甲因遇上好友多喝了酒，上公共汽车后呕吐，司机因躲避呕吐使车失控而打横，这时后面紧跟上来的另一辆车紧急转向道旁而翻入河中，造成惨案.

3. 偶然性的客观性

偶然性是客观的吗？这是历史上长期争论的著名问题，爱因斯坦与玻尔的争论实质上也与此有关. 一些人认为，世界是决定性的，偶然性只是出于人们的无知. 如果我们能预知一切情况，以后的发展便全已知，关于这点 1814 年拉普拉斯说得很明确：“智慧如果能在某一瞬间知道鼓动着自然的一切力量，知道大自然所有组成部分的相对位置，再者，如果它是如此浩瀚，足以分析这些材料，并能把上至宠大的天体、下至微小的原子的所有

运动悉数囊括于一个公式之中，那末，对于它来说，就没有什么东西是不可靠的了，无论是将来或过去，在它面前都会昭然若揭."(拉普拉斯：《概率论的哲学试验》) 按照这种观点，宇宙的一切发展，早在混沌初开时就已决定. 2 次世界大战、肯尼迪遇刺，某人手中所拿到的每付扑克牌的花色，都是百亿年前就已注定了的. 谁也不会同意这些意见，然而这竟出自大科学家之手笔，不能不令人深思：难道他毫无道理吗？

有些本来是必然的问题，由于无知，被人们当作随机问题来处理. 例如："存在地球外的生命吗？" 答案必然是 "是" 或 "否". 但现在，由于我们知识的贫乏，只能作出如下概率式的回答："存在球外生命的概率超过 90%". 类似地，人们常说："《金瓶梅》的作者很可能是王世贞","很有希望从地下挖出王羲之写的《兰亭序》" 等等. 对于这种人造的伪随机事件，拉普拉斯的观点是正确的.

对归一型试验，由于 $\lim\limits_{t \to e} p_k(t) = 1$, 它逐渐转变为必然试验. 因此，只要观察时刻落入转化区间，人们就能很准确地预言试验的结果. 对这种试验，拉普拉斯的论点也有充分理由.

然而，对于分散型试验，情况就不同了. 这里偶然性贯彻始终，直到试验结束，仍然至少有 2 个状态都可能出现. 无论知识如何丰富，决不能唯一地、决定性地预言试验的结果，预言只能是概率式的："出现 A_k 的概率为 p_k". 例如，无论怎么聪明，无论知识和经验如何丰富，谁也不能预知玩扑克时下次手中牌的花色. 由此可见，至少对分散型试验，随机性是客观存在的.

再考虑振动型. 不妨设只有 2 个状态 A_1 与 A_2(如果有 m 个，那么把 $A_3, \cdots, A_m$ 合并到 A_2 中成一新状态，仍记后者为 A_2). 由振动型定义，存在 2 个不相交的数列 $s_i \to e$, $t_i \to e$, 使当 $i \to \infty$ 时有

$$p_1(s_i) \to a = \overline{\lim_{t \to e}} p_1(t), \quad p_1(t_i) \to b = \underline{\lim_{t \to e}} p_1(t),$$

其中 $1 \geqslant a > b \geqslant 0$. 由于 $p_1(t) + p_2(t) = 1$, 故

$$p_2(s_i) \to 1 - a, \quad p_2(t_i) \to 1 - b.$$

共有 3 种情形：

a. $a = 1, b = 0$ 于是

$$p_1(s_i) \to 1, \quad p_2(t_i) \to 1.$$

从而沿 $\{s_i\}$, A_1, 逐步变为必然状态，而沿 $\{t_i\}$, 则 A_2 变为必然的. 于是在此二时间序列 $\{s_i\}, \{t_i\}$ 上，此试验分别化为二不同的归一型试验，其一最终必出现 A_1, 另一最终必出现 A_2.

b. $a < 1$ 这时

$$p_1(s_i) \to a > 0, \quad p_2(s_i) \to 1 - a > 0,$$

于是沿 $\{s_i\}$, 此试验是分散型的.

c. $b > 0$ 这时

$$p_1(t_i) \to b > 0, \quad p_2(t_i) \to 1 - b > 0,$$

于是沿 $\{t_i\}$, 此试验也是分散型的.

复合试验，如果不是人的预谋，大多数是不能准确预测的. 因为，只要其中包含一个分散型的子试验，整个试验的结果便不能确切地唯一地决定.

4. 非重复随机试验

科学需要重复. 不能重复的一次性现象，科学中一般不予研究，因为缺乏客观的检验. 然而实际中有许多问题，意义非常重大，人们津津有味地议论它们，却不幸是非重复的. 例如：近期能再爆发世界大战吗？某人会患癌症吗？某地会发生大地震吗？我们能否计算这些事件的概率？

原则上说，概率由试验的条件决定；只要给出了随机试验，概率也就决定了. 但在实际问题中，情况要复杂得多，一是试验的条件有时难以确切地叙述，例如，把爆发大战看成试验，条件应该是那些呢？二是从条件推算出概率，也往往很不容易，演绎推理或计算，有时非常困难.

对于重复试验，如果精度要求不很高，总可以用频率逼近以求出概率的近似值. 但对非重复试验，已无频率可言；那么，有无其他近似方法呢？

一种方法是适当修改条件. 设试验 E 的条件组为 C, 在 C 下 E 非重复；今适当修改 C 为 C', 而在 C' 下则是可重复的，于是化归重复试验，其代价是要牺牲一些精度. 采用此法时，还要求试验结果对条件不要过于敏感，即要求试验有适度的稳定性，以免开始条件“差之毫厘”, 而结果却“失之千里”, 出现所谓“混沌现象”.

例如，试求“地球以外存在生命”的概率. 本来，每颗星球的情况千差万别，但可采用下述近似方法. 限于目前科学水平，只考虑银河系. 银河系中恒星数约为 3×10^{11}, 其中生物可居住的行星数约为 65×10^{7}. 这些行星中，设每颗有生命的概率不小于 ε; 于是每颗星无生命之概率不大于 $1-\varepsilon$; 于是 65×10^{7} 颗行星都无生命之概率不大于 $(1-\varepsilon)^{65\times10^{7}}$; 于是至少还有一颗行星上有生命的概率不小于 $1-(1-\varepsilon)^{65\times10^{7}}$. 不管 ε 如何小，只要它大于 0, 最后一概率充分接近于 1, 因此，我们对存在球外生命甚有信心. 在这里，我们把观察 1 颗行星当作 1 次试验，并把它重复 65×10^{7} 次.

又如，要求出病人李某存活期的分布. 通过检查，发现他的病情与其他一些人相似，而后者皆已作古，其存活期是有案可查的. 根据记录可制出已作古者存活期的经验分布，于是人们便以此分布作为李某存活期分布，从而预测他还能活多少时间及对应的概率. 其实李某未必与那些人情况全同，只是大体相似罢了.

保险公司也可采用其他方法，例如因素评分法. 列出影响寿命的一些重要因素如心脏、血管、大脑的病化程度等；把每因素的指标分成若干等级；给每一等级评分 (相当于“加权”); 最后根据李某所得总分按照某种算法以求出其存活期的分布. 此法既有一定的科学道理，但也有主观性，因为评分多少以及采用什么算法都有选择的随意性. 对此人们不必大惊小怪，其实许多社会活动如选拔，比赛，评估等等都离不开因素评分法，它常常能在很大程度上反映实际情况.

除上述 2 种方法外，还可采用模拟方法以计算非重复事件的概率.

参考文献

1. 田景春，古生物大规模绝灭假说种种，世界科学， 1990, 2:37.

2. Poincare H, 机遇，数学译林， 1987, 6(2):165.

3. 王梓坤，论混沌与随机，北京师范大学学报 (自然科学版), 30 卷 2 期， 1994, 199-202.

参 考 书 目

[1] И. П. Натансон, 实变函数论，第二版， 1957(徐瑞云译).

[2] M. Loéve, Probability Theory, 第三版， 1963.

[3] H. Cramer, Mathematical Methods of Statistics, 1946 (有汉译本；魏宗舒译，统计学数学方法).

[4] Н. И. Ахиезер, Классическая Проблема Моментов, 1961.

[5] E. Lukacs, Characteristic Function, 1960.

[6] S. S. Wilk, Mathematical Statistics, 1962.

[7] M. Fisz, 概率论及数理统计， 1958(王福保译).

[8] 复旦大学数学系主编，概率论与数理统计，第二版， 1961.

[9] Б. В. Гнеденко, 概率论教程，第二版， 1954(丁寿田译).

[10] 江泽坚、吴智泉，实变函数论，第二版.

[11] Б. В. Гнеденко, А. Н. Колмогоров, 相互独立随机变数之和的极限分布， 1949(王寿仁译).

[12] H. Richter, Wahrscheinlichkeitstheorie, 1956.

[13] Ю. В. Линник, Разложния Вероятностных Законов, 1960.

[14] H. Cramer, Random Variables and Probability Distributions, 1937

[15] W. Feller, An Introduction to Probability Theory and Its Applications, Vol.1, 第二版， 1957; Vol.2, 第二版， 1971.

[16] J. L. Doob, Stochastic Processes, 1953.

[17] 伊藤清，概率论， 1952(刘璋温译).

[18] 伊藤清，随机过程 (刘璋温译， 1961).

[19] A. Blanc-Lapierre, R. Fortet, Théoric des Fonction Aléatoires, 1953.

[20] M. S. Bartlett, An Introduction to Stochastic Processes, 1955.

[21] A. T. Bharucha-Reid, Elements of the Theory of Markov Processes and Their Applications, 1960.

[22] 雅格龙，平稳随机函数导论，数学进展， 2,1(1956),3-153.

[23] Ю. А. Розанов, Стационарные Случайные Процессы,1963.

[24] 郑绍濂编校，希尔伯脱空间中的平稳序列， 1963.

[25] N. Grenander, M. Rosenblatt, Statistical Analysis of Stationary Time Series, 1957(有汉译本，郑绍濂译).

[26] Е. Б. Дынкин, 马尔科夫过程论基础， 1959(王梓坤译).

[27] Е. Б. Дынкин, Марковские Процессы, 1963.

[28] А. В. Скороход, Исследования по Теории Случайных Процессов.

[29] K. L. Chung, Markov Chains with Stationary Transition Probablities, 1960.

[30] И. С. Вентцепь 概率论，第二版.

[31] M. G. Kendall, The Advanced Theory of Statistics, Fifth edition, Vol. 1,2, 1952.

[32] 周华章，工业技术应用数理统计学， 1963.

[33] И. В. Дунин-Барковский, Н. В. Смироов, Теория Вероятносгей и Математическая Статистика в Технике, 1955.

[34] Р. М. Фихтенгольц, 微积分学教程 (汉译本， 1964).

[35] 王梓坤，随机过程论， 1965.
[36] A. Wald, Sequential Analysis, 1948.
[37] Д. И. Голенко, Моделирование и Статистический Анализ Псевдослучайных Числа на Электронных Вычислительных Машинах, 1965.
[38] И. А. Ибрагимов, Ю. В. Линник, Независимые и Стационарно Связанные Величины, 1965.
[39] E. Parzen, Stochastic Processes, 1962.
[40] Б. П. Демидович и И. А. Марон, Основы Вычислительной Математики, 1960.
[41] 林少宫，基础概率与数理统计， 1964.
[42] 中国科学院数学研究所统计组编：常用数理统计方法， 1973.
[43] 王宗皓，李麦村等，天气预报中的概率统计方法， 1974.
[44] R. E. Barlow 等， Mathematical Theory of Realiability, 1965.
[45] Б. В. Гнеденко 等， Математические Методы в Теории Надежности, 1965.
[46] M. Kac, Probability and Related Topics in Physical Sciences, 1959.
[47] 中国科学院数学研究所概率统计室，常用数理统计表， 1974.
[48] 中国科学院数学研究所概率统计室，回归分析方法， 1974.
[49] В. В. Петров, Суммы Независимых Случайных Величин, 1972.
[50] H. Cramer, The Elements of Probability Theory and Some of its Applications, 1955.
[51] 杨纪珂，数理统计方法在医学科学中的应用， 1964.
[52] 严士健，王隽骧，刘秀芳，概率论基础，科学出版社， 1982.
[53] 严士健，刘秀芳，徐承彝，概率论与数理统计，高等教育出版社， 1990.
[54] 汪嘉冈，现代概率论基础，复旦大学出版社， 1988.
[55] 杨宗磐，概率论入门，科学出版社， 1981.
[56] 林叔荣，实用统计决策与 Bayes 分析，厦门大学出版社， 1991.
[57] 林鸿庆，概率论，厦门大学出版社， 1989.
[58] 陈希孺，倪国熙，数理统计教程，上海科技出版社， 1988.
[59] 钟开莱，概率论教程，上海科学技术出版社， 1989.
[60] 钟开莱，初等概率论附随机过程，人民教育出版社， 1980.
[61] 胡迪鹤，分析概率论，科学出版社， 1984.
[62] 梁之舜等，概率论与数理统计，高等教育出版社， 1992.
[63] 廖昭懋，杨文礼，概率论与数理统计，北京师范大学出版社， 1988.
[64] 陆传荣，林正炎，陆传赉，概率论极限理论引论.
[65] Yuan Shih Chow, Henry Teicher, Probability Theory, Independence, Interchangeability, Martingales, Second edition, Springer-Verlag, World Publishing Corporation, 1989.
[66] 朱成熹，随机极限引论，南开大学出版社， 1989.
[67] 周概容，概率论与数理统计，高等教育出版社， 1984.
[68] 骆振华，概率统计简明教程，厦门大学出版社， 1990.
[69] 谢尔登・罗斯，概率论初级教程，李漳南，杨振明译，人民教育出版社， 1982.
[70] 魏宗舒等，概率论与数理统计教程，高等教育出版社， 1991.
[71] 王梓坤，生灭过程与马尔科夫链，科学出版社， 1980.

[72] 王梓坤，布朗运动与位势，科学出版社， 1983.
[73] 王梓坤，钱尚玮，概率与统计预报，科学出版社， 1978.

内 容 索 引

十二画

十三画

十三画以上

外来字

[General Information]
书名=概率论基础及其应用
作者=王梓坤
页数=351
SS号=11513929
DX号=
出版日期=1996年05月第1版
出版社=北京师范大学出版社